Adaptive procedures in convolution models with known or partially known noise distribution

Cristina Butucea, Catherine Matias, Christophe Pouet

Adaptive procedures in convolution models with known or partially known noise distribution

Cristina Butucea [a,b]

[a] *Laboratoire de Probabilités et Modèles Aléatoires (UMR CNRS 7599), Université Paris VI, 4, pl.Jussieu, Boîte courrier 188, 75252 Paris, France*

[b] *Modal'X, Université Paris X, 200, avenue de la République 92001 Nanterre Cedex, France*

Catherine Matias [c,∗]

[c] *Laboratoire Statistique et Génome (UMR CNRS 8071), Tour Evry 2, 523 pl. des Terrasses de l'Agora, 91 000 Evry, France*

Christophe Pouet [d]

[d] *Laboratoire d'Analyse, Topologie, Probabilités (UMR CNRS 6632), Centre de Mathématiques et Informatique, Université de Provence, 39 rue F. Joliot-Curie, 13453 Marseille cedex 13, France*

Abstract

In a convolution model, we observe random variables whose distribution is the convolution of some unknown density f and some known or partially known noise density g. In this paper, we focus on statistical procedures, which are adaptive with respect to the smoothness parameter τ of unknown density f, and also (in some cases) to some unknown parameter of the noise density g.

In a first part, we assume that g is known and polynomially smooth. We provide goodness-of-fit procedures for the test $H_0 : f = f_0$, where the alternative H_1 is expressed with respect to $\mathbb{L}_2$-norm (i.e. has the form $\psi_n^{-2}\|f - f_0\|_2^2 \geq \mathcal{C}$). Our adaptive (w.r.t τ) procedure behaves differently according to whether f_0 is polynomially or exponentially smooth. A payment for adaptation is noted in both cases and for computing this, we provide a non-uniform Berry-Esseen type theorem for degenerate U-statistics. In the first case we prove that the payment for adaptation is optimal (thus unavoidable).

In a second part, we study a wider framework: a semiparametric model, where g is exponentially smooth and stable, and its self-similarity index s is unknown. In order to ensure identifiability, we restrict our attention to polynomially smooth, Sobolev-type densities f. In this context, we provide a consistent estimation procedure for

s. This estimator is then plugged-into three different procedures: estimation of the unknown density f, of the functional $\int f^2$ and test of the hypothesis H_0. These procedures are adaptive with respect to both s and τ and attain the rates which are known optimal for known values of s and τ. As a by-product, when the noise is known and exponentially smooth our testing procedure is adaptive for testing Sobolev-type densities.

Résumé

Dans un modèle de convolution, les observations sont des variables aléatoires réelles dont la distribution est la convoluée entre une densité inconnue f et une densité de bruit g supposée soit entièrement connue, soit connue seulement à paramètre près. Nous étudions différentes procédures statistiques qui s'adaptent automatiquement au paramètre de régularité τ de la densité inconnue f ainsi que (dans certains cas), au paramètre inconnu de la densité du bruit.

Dans une première partie, nous supposons que g est connue et de régularité polynomiale. Nous proposons un test d'adéquation de l'hypothèse $H_0 : f = f_0$ lorsque l'alternative H_1 est exprimée à partir de la norme $\mathbb{L}_2$ (i.e. de la forme $\psi_n^{-2}\|f - f_0\|_2^2 \geq \mathcal{C}$). Cette procédure est adaptative (par rapport à τ) et présente différentes vitesses de test (ψ_n) en fonction du type de régularité de f_0 (polynomiale ou bien exponentielle). L'adaptativité induit une perte sur la vitesse de test, perte qui est calculée grâce à un théorème de type Berry-Esseen non-uniforme pour des U-statistiques dégénérées. Dans le cas d'une régularité polynomiale pour f, nous prouvons que cette perte est inévitable et donc optimale.

Dans un second temps, nous nous placons dans le cadre plus large d'un modèle semi-paramétrique, où g est la densité d'une loi stable (régularité de type exponentiel) avec un indice d'auto-similarité s inconnu. Pour assurer l'identifiabilité du modèle, la densité f est supposée appartenir à un espace de Sobolev (régularité polynomiale). Dans ce cadre, nous proposons un estimateur consistant de s. Celui-ci est ensuite injecté dans trois procédures différentes : l'estimation de f, de la fonctionnelle $\int f^2$ et le test de l'hypothèse H_0. Ces procédures sont adaptatives par rapport à s et à τ et atteignent les vitesses optimales du cas s et τ connus. Enfin, lorsque g est connue et de régularité exponentielle, une conséquence de notre résultat est que cette procédure de test est adaptative lorsque f_0 appartient à un espace de Sobolev.

Key words: Adaptive nonparametric tests, convolution model, goodness-of-fit tests, infinitely differentiable functions, partially known noise, quadratic functional estimation, Sobolev classes, stable laws
PACS: 62F12, 62G05, 62G10, 62G20

* Corresponding author
Email addresses: `butucea@ccr.jussieu.fr` (Cristina Butucea), `matias@genopole.cnrs.fr` (Catherine Matias), `pouet@cmi.univ-mrs.fr` (Christophe Pouet).

1 Introduction

Convolution model

Consider the **convolution model** where the observed sample $\{Y_j\}_{1\le j\le n}$ comes from the independent sum of independent and identically distributed (i.i.d.) random variables X_j with unknown density f and Fourier transform Φ and i.i.d. noise variables ε_j with known (maybe only up to a parameter) density g and Fourier transform Φ^g

$$Y_j = X_j + \varepsilon_j, \quad 1 \le j \le n. \tag{1}$$

The density of the observations is denoted by p and its Fourier transform Φ^p. Note that we have $p = f * g$ where $*$ denotes the convolution product and $\Phi^p = \Phi\Phi^g$.

The underlying unknown density f is always supposed to belong to $\mathbb{L}_1 \cap \mathbb{L}_2$. We shall consider probability density functions belonging to the class

$$\mathcal{F}(\alpha, r, \beta, L) = \left\{ f : \mathbb{R} \to \mathbb{R}_+, \int f = 1, \frac{1}{2\pi} \int |\Phi(u)|^2 |u|^{2\beta} \exp(2\alpha |u|^r)\, du \le L \right\}, \tag{2}$$

for L a positive constant, $\alpha > 0$, $0 \le r \le 2$, $\beta \ge 0$ and either $r > 0$ or $\beta > 0$. Note that the case $r = 0$ corresponds to Sobolev densities whereas $r > 0$ corresponds to infinitely many differentiable (or supersmooth) densities.

We consider noise distributions whose Fourier transform does not vanish on $\mathbb{R}$: $\Phi^g(u) \ne 0$, $\forall\ u \in \mathbb{R}$. Typically, nonparametric estimation in convolution models gives rise to the distinction of two different behaviours for the noise distribution. We alternatively shall consider (for some constant $c_g > 0$),

polynomially smooth (or polynomial) noise

$$|\Phi^g(u)| \sim c_g |u|^{-\sigma}, \ |u| \to \infty, \sigma > 1; \tag{3}$$

exponentially smooth (or supersmooth or exponential) **stable** noise

$$|\Phi^g(u)| = \exp(-\gamma |u|^s), \ |u| \to \infty, \gamma, s > 0. \tag{4}$$

In this second case, the parameter s is called the self-similarity index of the noise density and we shall consider that it is **unknown**.

Convolution models have been widely studied over the past two decades. We will be interested here both in estimation of the unknown density f and in testing the hypothesis $H_0 : f = f_0$, with a particular interest in *adaptive* procedures. Our first purpose is to provide goodness-of-fit testing procedures on f, for the test of the hypothesis $H_0 : f = f_0$, which are adaptive with respect to the unknown smoothness parameter of f. The second one is to study the behaviour of different procedures (such as estimation of f, estimation of $\int f^2$ and goodness-of-fit test) in a setup where self-similarity index s is unknown.

Adaptive procedures in the convolution model

Concerning estimation, the asymptotically minimax setup in the context of pointwise or $\mathbb{L}_p$-norms and in the case of entirely known noise density g is the most studied one. Major results in this direction prove that the smoother the error density, the slower the optimal rates of convergence (see [6], [9], [2], [10] concerning polynomial noise and [21], [7], [5] for exponential noise). Adaptive estimation procedures were considered first by [21] and then by [11]. They constructed wavelets estimators which do not depend on smoothness parameter of the density f to be estimated. Adaptive kernel estimators were given in [5]. A different adaptive approach is used in [7] relying on penalized contrast estimators.

Nonparametric goodness-of-fit testing has extensively been studied in the context of direct observations (namely a sample distributed from the density f to be tested), but also for regression or in the Gaussian white noise model. We refer to [18], [16] for an overview on the subject. The convolution model provides an interesting setup where observations may come from a signal observed through some noise.

Nonparametric goodness-of-fit tests in convolution models were studied in [15] and in [3]. The approach used in [3] is based on a minimax point of view combined with estimation of the quadratic functional $\int f^2$. Assuming the smoothness parameter of f to be known, the authors of [15] define a version of the Bickel-Rosenblatt test statistic and study its asymptotic distribution under the null hypothesis and under fixed and local alternatives, while [3] provides a different goodness-of-fit testing procedure attaining the minimax rate of testing in each of the three following setups: Sobolev densities and polynomial noise, supersmooth densities and polynomial noise, Sobolev densities and exponential noise. The case of supersmooth densities and exponential noise is also studied but the optimality of the procedure is not established in the case $r > s$.

Our first goal here is to provide adaptive versions of these last procedures with respect to the parameters (α, r, β). We restrict our attention to testing problems where alternatives are expressed with respect to $\mathbb{L}_2$-norm. Namely, the alternative has the form $H_1 : \psi_n^{-2}\|f - f_0\|_2^2 \geq \mathcal{C}$. In such a case, the problem relates with asymptotically minimax estimation of $\int f^2$.

Our second goal is to deal with the case of not entirely known noise distribution. This is a crucial issue as, assuming this noise distribution to be entirely known is not realistic in many situations. However, in general, the noise density g has to be known for the model to be identifiable. Nevertheless, when the noise density is exponentially smooth and the unknown density is restricted to be less smooth than the noise, semiparametric models are identifiable and they may be considered. The case of a Gaussian noise with unknown variance γ and unknown density without Gaussian component has first been considered in [19]. She proposes an estimator of the parameter γ which is then plugged in an estimator of the unknown density. This work is generalized in [4] for exponentially smooth noise with unknown scale parameter γ and unknown densities belonging either to Sobolev classes, or to classes of supersmooth densities with parameter r, $r < s$. Minimax rates of convergence are exhibited. In this context, the unknown parameter γ acts as a real nuisance parameter as the rates of convergence for estimating the unknown density are slower compared to the case of known scale, those rates being nonetheless optimal in a minimax sense. Another attempt to remove knowledge on the noise density appears in [20] where the author studies a deconvolution estimator associated to a procedure for selecting the error density between the Normal supersmooth density and the Laplace polynomially smooth density (both with fixed parameter values).

In the second part of our work, we will be interested in estimation procedures on f, adaptive both with respect to the smoothness parameter of f and to an unknown parameter of the noise density. More precisely, in the specific setup of Sobolev densities and exponential noise with symmetric stable distribution, we will consider the case of unknown self-similarity index s. In this context, we first propose an estimator of the self-similarity index s, which, plugged into kernel procedures, provides estimators of the unknown density f with the same optimal rate of convergence as in the case of entirely known noise density. Using the same techniques, we also construct an estimator of the quadratic functional $\int f^2$ (with optimal rate of convergence) and $\mathbb{L}_2$ goodness-of-fit test statistic. Note that this work is very different from [4] as the self similarity index s plays a different role from the scale parameter γ previously studied. In particular, the range of applications of those results is entirely new.

In the sequel, $\|\cdot\|_2$ denotes the $\mathbb{L}_2$-norm, $\bar{M}$ is the complex conjugate of M and $< M, N >= \int M(x)\bar{N}(x)dx$ is the scalar product of complex-valued functions in $\mathbb{L}_2(\mathbb{R})$. Moreover, probability and expectation with respect to the distribution of $Y_1, \ldots, Y_n$ induced by the unknown density f will be denoted by $\mathbb{P}_f$ and $\mathbb{E}_f$.

We denote more generally by $\tau = (\alpha, r, \beta)$ the smoothness parameter of the unknown density f and by $\mathcal{F}(\tau, L)$ the corresponding class. As the density f is unknown, the a priori knowledge of its smoothness parameter τ could appear unrealistic. Thus, we assume that τ belongs to a closed subset $\mathcal{T}$, included in $(0, +\infty) \times (0, 2] \times (0, +\infty)$. For a given density f_0 in the class $\mathcal{F}(\tau_0)$, we want to test the hypothesis

$$H_0 : f = f_0$$

from observations $Y_1, \ldots, Y_n$ given by (1). We extend the results of [3] by giving the family of sequences $\Psi_n = \{\psi_{n,\tau}\}_{\tau \in \mathcal{T}}$ which separates (with respect to $\mathbb{L}_2$-norm) the null hypothesis from a larger alternative

$$H_1(\mathcal{C}, \Psi_n) : f \in \cup_{\tau \in \mathcal{T}} \{f \in \mathcal{F}(\tau, L) \text{ and } \psi_{n,\tau}^{-2} \|f - f_0\|_2^2 \geq \mathcal{C}\}.$$

We recall that the usual procedure is to construct, for any $0 < \epsilon < 1$, a test statistic $\Delta_n^\star$ (an arbitrary function, with values in $\{0, 1\}$, which is measurable with respect to $Y_1, \ldots, Y_n$ and such that we accept H_0 if $\Delta_n^\star = 0$ and reject it otherwise) for which there exists some $\mathcal{C}^0 > 0$ such that

$$\limsup_{n\to\infty} \left\{ \mathbb{P}_0[\Delta_n^\star = 1] + \sup_{f \in H_1(\mathcal{C}, \Psi_n)} \mathbb{P}_f[\Delta_n^\star = 0] \right\} \leq \epsilon, \tag{5}$$

holds for all $\mathcal{C} > \mathcal{C}^0$. This part is called the upper bound of the testing rate. Then, prove the minimax optimality of this procedure, i.e. the lower bound

$$\liminf_{n\to\infty} \inf_{\Delta_n} \left\{ \mathbb{P}_0[\Delta_n = 1] + \sup_{f \in H_1(\mathcal{C}, \Psi_n)} \mathbb{P}_f[\Delta_n = 0] \right\} \geq \epsilon, \tag{6}$$

for some $\mathcal{C}_0 > 0$ and for all $0 < \mathcal{C} < \mathcal{C}_0$, where the infimum is taken over all test statistics Δ_n.

Let us first remark that as we use noisy observations (and unlike what happens with direct observations), this test cannot be reduced to testing uniformity of the distribution density of the observed sample (i.e. $f_0 = 1$ with support on the finite interval $[0; 1]$). As a consequence, additional assumptions used in [3] on the tail behaviour of f_0 (ensuring it does not vanish arbitrarily fast) are needed to obtain the optimality result of the testing procedure in the case of Sobolev density ($r = 0$) observed with polynomial noise ((**T**) and (**P**)), respectively with exponential noise ((**T**) and (**E**)). We recall these assumptions here for reader's convenience.

Assumption (T)

$$\exists c_0 > 0, \forall x \in \mathbb{R}, f_0(x) \geq \frac{c_0}{1 + |x|^2}.$$

Moreover, we also need to control the derivatives of known Fourier transform Φ^g when establishing optimality results.

Assumption (P) (Polynomial noise) If the noise satisfies (3), then assume that Φ^g is three times continuously differentiable and there exist A_1, A_2 such that

$$|(\Phi^g)'(u)| \leq \frac{A_1}{|u|^{\sigma+1}} \text{ and } |(\Phi^g)''(u)| \leq \frac{A_2}{|u|^{\sigma+2}}, \quad |u| \to \infty.$$

Assumption (E) (Exponential noise) If the noise satisfies (4), then assume that Φ^g is continuously differentiable and there exists some constants $C > 0$ and $A_3 \in \mathbb{R}$ such that

$$|(\Phi^g)'(u)| \leq C|u|^{A_3} \exp(-\gamma|u|^s), \quad |u| \to \infty.$$

Remark 1 *Similar results may be obtained when we assume the existence of some $p \geq 1$ such that $f_0(x)$ is bounded from below by $c_0(1 + |x|^p)^{-2}$ for large enough x. In such a case, the Fourier transform Φ^g of the noise density is assumed to be p times continuously differentiable, with derivatives up to order p satisfying the same kind of bounds as in Assumption* (**P**)*, when the noise is polynomial, respectively in Assumption* (**E**)*, when the noise is exponential.*

Roadmap

Section 2 deals with the case of (known) polynomial noise. We provide a goodness-of-fit testing procedure for the test $H_0 : f = f_0$, in two different cases: the density f_0 to be tested is either ordinary smooth ($r_0 = 0$) or supersmooth ($r_0 > 0$). The procedures are adaptive with respect to the smoothness parameter (α, r, β) of f. The proof of the upper bounds for the testing rate relies mainly on a Berry-Esseen inequality for degenerate U-statistics of order 2, postponed to Section 4. In some cases, a loss for adaptation is noted with respect to known testing rates for fixed known parameters. When the loss is of order $\log\log n$ to some power, we prove that this payment is unavoidable.

In Section 3, we consider exponential noise of symmetric stable law with unknown self-similarity index s. In order to ensure identifiability, we restrict our attention to Sobolev classes of densities f. The first step (Section 3.1) is to provide a consistent estimation procedure for the self-similarity index. Then (Section 3.2) using a plug-in, we introduce a new kernel estimator of f where both the bandwidth and the kernel are data dependent. We also introduce an estimator of the quadratic functional $\int f^2$ with sample dependent bandwidth and kernel. We prove that these two procedures attain the same rates of convergence as in the case of entirely known noise distribution, and are thus asymptotically optimal in the minimax sense. We also present a goodness-of-fit test on f in this setup. We prove that the testing rate is the same as in the case of entirely known noise distribution and thus asymptotically optimal in the minimax sense. Proofs are postponed to Section 5.

2 Polynomially smooth noise

In this section, we shall assume that the noise density g is polynomial (3). The unknown density f belongs to the class $\mathcal{F}(\alpha, r, \beta, L)$. We are interested in adaptive, with respect to the parameter $\tau = (\alpha, r, \beta)$, goodness-of-fit testing procedures. We assume that this unknown parameter belongs to the following set

$$\mathcal{T} = \{\tau = (\alpha, r, \beta); \tau \in [\underline{\alpha}; +\infty) \times [\underline{r}; \bar{r}] \times [\underline{\beta}; \bar{\beta}]\},$$

where $\underline{\alpha} > 0$, $0 \leq \underline{r} \leq \bar{r} \leq 2$, $0 \leq \underline{\beta} \leq \bar{\beta}$ and either $\underline{r} > 0$ and $\alpha \in [\underline{\alpha}, \overline{\alpha}]$ or both $\underline{r} = \bar{r} = 0$ and $\underline{\beta} > 0$.

Let us introduce some notation. We consider a preliminary kernel J, with Fourier transform Φ^J, defined by

$$\forall x \in \mathbb{R}, \quad J(x) = \frac{\sin(x)}{\pi x}, \qquad \forall u \in \mathbb{R}, \quad \Phi^J(u) = 1_{|u| \leq 1},$$

where 1_A is the indicator function of the set A. For any bandwidth $h = h_n \to 0$ as n tends to infinity, we define the rescaled kernel J_h by

$$\forall x \in \mathbb{R}, \quad J_h(x) = h^{-1} J(x/h) \text{ and } \forall u \in \mathbb{R}, \quad \Phi^{J_h}(u) = \Phi^J(hu) = 1_{|u| \leq 1/h}.$$

Now, the deconvolution kernel K_h with bandwidth h is defined via its Fourier transform Φ^{K_h} as

$$\Phi^{K_h}(u) = (\Phi^g(u))^{-1} \Phi^J(uh) = (\Phi^g(u))^{-1} \Phi^{J_h}(u), \quad \forall u \in \mathbb{R}. \tag{7}$$

In Section 3.2, we will consider a modification of this kernel to take into account the case of not entirely known noise density g.

Next, the quadratic functional $\int (f - f_0)^2$ is estimated by the statistic $T_{n,h}$

$$T_{n,h} = \frac{2}{n(n-1)} \sum\sum_{1 \le k < j \le n} < K_h(\cdot - Y_k) - f_0, K_h(\cdot - Y_j) - f_0 > . \quad (8)$$

Note that $T_{n,h}$ may not be positive, but its expected value is.

In order to construct a testing procedure which is adaptive with respect to the parameter τ we introduce a sequence of finite regular grids over the set $\mathcal{T}$ of unknown parameters: $\mathcal{T}_N = \{\tau_i; 1 \le i \le N\}$. For each grid point τ_i we choose a testing threshold $t_{n,i}^2$ and a bandwidth h_n^i giving a test statistic T_{n,h_n^i}.

The test rejects the null hypothesis as soon as at least one of the single tests based on the parameter τ_i is rejected.

$$\Delta_n^\star = \begin{cases} 1 \text{ if } \sup_{1 \le i \le N} |T_{n,h_n^i}| t_{n,i}^{-2} > \mathcal{C}^\star \\ 0 \text{ otherwise,} \end{cases} \quad (9)$$

for some constant $\mathcal{C}^\star > 0$ and finite sequences of bandwidths $\{h_n^i\}_{1 \le i \le N}$ and thresholds $\{t_{n,i}^2\}_{1 \le i \le N}$.

We note that our asymptotic results work for large enough constant $\mathcal{C}^\star$. In practice we may choose it by Monte-Carlo simulation under the null hypothesis, for known f_0, such that we control the first-type error of the test and bound it from above, e.g. by $\epsilon/2$.

Typically, the structure of the grid accounts for two different phenomena. A first part of the points is dedicated to the adaptation with respect to β in case $\bar{r} = \underline{r} = 0$, whereas the rest of the points is used to adapt the procedure with respect to r (whatever the value of β).

In the two next theorems, we fix $\sigma > 1$. We note that the testing rates are essentially different according to the two different cases where f_0 belongs to a Sobolev class ($r_0 = 0$, $\alpha_0 \ge \underline{\alpha}$ and we assume $\beta_0 = \bar{\beta}$) and where f_0 is a supersmooth function ($\alpha_0 \in [\underline{\alpha}, \overline{\alpha}]$, $r_0 > 0$ and $\beta_0 \in [\underline{\beta}, \bar{\beta}]$ and then we focus on $r_0 = \bar{r}$ and $\alpha_0 = \overline{\alpha}$). Note that in the first case, the alternative contains functions f which are smoother ($r > 0$) than the null hypothesis f_0.

When f_0 belongs to Sobolev class $\mathcal{F}(\alpha_0, 0, \bar{\beta}, L)$, the grid is defined as follows. Let N and choose $\mathcal{T}_N = \{\tau_i; 1 \le i \le N+1\}$ such that

$$\begin{cases} \forall 1 \le i \le N, \tau_i = (0; 0; \beta_i) \text{ and } \beta_1 = \underline{\beta} < \beta_2 < \ldots < \beta_N = \bar{\beta}, \\ \forall 1 \le i \le N-1,\ \beta_{i+1} - \beta_i = (\bar{\beta} - \underline{\beta})/(N-1), \\ \text{ and } \tau_{N+1} = (\underline{\alpha}; \bar{r}; 0) \end{cases}$$

In this case, the first N points are dedicated to the adaptation with respect to β when $\bar{r} = \underline{r} = 0$, whereas the last point τ_{N+1} is used to adapt the procedure with respect to r (whatever the value of β).

Theorem 1 *Assume $f_0 \in \mathcal{F}(\alpha_0, 0, \bar{\beta}, L)$. The test statistic $\Delta_n^\star$ given by (9) with parameters*

$$N = \lceil \log n \rceil; \quad \forall 1 \le i \le N : \begin{cases} h_n^i = \left(\frac{n}{\sqrt{\log \log n}} \right)^{-2/(4\beta_i + 4\sigma + 1)} \\ t_{n,i}^2 = \left(\frac{n}{\sqrt{\log \log n}} \right)^{-4\beta_i/(4\beta_i + 4\sigma + 1)} \end{cases},$$
$$h_n^{N+1} = n^{-2/(4\bar{\beta} + 4\sigma + 1)}; \quad t_{n,N+1}^2 = n^{-4\bar{\beta}/(4\bar{\beta}+4\sigma+1)},$$

and any large enough positive constant $\mathcal{C}^\star$, satisfies (5) *for any $\epsilon \in (0,1)$, with testing rate $\Psi_n = \{\psi_{n,\tau}\}_{\tau \in \mathcal{T}}$ given by*

$$\psi_{n,\tau} = \left(\frac{n}{\sqrt{\log \log n}} \right)^{-2\beta/(4\beta+4\sigma+1)} 1_{r=0} + n^{-2\bar{\beta}/(4\bar{\beta}+4\sigma+1)} 1_{r>0}, \; \forall \tau = (\alpha, r, \beta) \in \mathcal{T}.$$

Moreover, if $f_0 \in \mathcal{F}(\alpha_0, 0, \bar{\beta}, cL)$ for some $0 < c < 1$ and if Assumptions **(T)** *and* **(P)** *hold, then this testing rate is adaptive minimax over the family of classes $\{\mathcal{F}(\tau, L), \tau \in [\underline{\alpha}, \infty) \times \{0\} \times [\underline{\beta}, \bar{\beta}]\}$ (i.e.* (6) *holds).*

We note that our testing procedure attains the polynomial rate $n^{-2\bar{\beta}/(4\bar{\beta}+4\sigma+1)}$ over the union of all classes containing functions smoother than f_0. Note moreover that this rate is known to be a minimax testing rate over the class $\mathcal{F}(0, 0, \bar{\beta}, L)$ by results in [3]. Therefore we prove that the loss of some power of $\log \log n$ with respect to the minimax rate is unavoidable. A loss appears when the alternative contains classes of functions less smooth than f_0.

The proof that our adaptive procedure attains the minimax rate relies on the Berry-Esseen inequality presented in Section 4.

When f_0 belongs to class $\mathcal{F}(\overline{\alpha}, \bar{r}, \beta_0, L)$ of infinitely many differentiable functions, the grid is defined as follows. Let N_1, N_2 and choose $\mathcal{T}_N = \{\tau_i; 1 \le i \le N = N_1 + N_2\}$ such that

$$\begin{cases} \forall 1 \le i \le N_1, \tau_i = (0; 0; \beta_i) \text{ and } \beta_1 = \underline{\beta} < \beta_2 < \ldots < \beta_{N_1} = \bar{\beta}, \\ \forall 1 \le i \le N_1 - 1, \; \beta_{i+1} - \beta_i = (\bar{\beta} - \underline{\beta})/(N_1 - 1), \\ \text{ and } \forall 1 \le i \le N_2, \tau_{N_1+i} = (\overline{\alpha}; r_i; \beta_0) \text{ and } r_1 = \underline{r} < r_2 < \ldots < r_{N_2} = \bar{r}, \\ \forall 1 \le i \le N_2 - 1, \; r_{i+1} - r_i = (\bar{r} - \underline{r})/(N_2 - 1). \end{cases}$$

In this case, the first N_1 points are used for adaptation with respect to β in case $\bar{r} = \underline{r} = 0$, whereas the last N_2 points are used to adapt the procedure with respect to r (whatever the value of β).

Theorem 2 *Assume $f_0 \in \mathcal{F}(\overline{\alpha}, \bar{r}, \beta_0, L)$ for some $\beta_0 \in [\underline{\beta}, \bar{\beta}]$. The test statistic $\Delta_n^\star$ given by* (9) *with $\mathcal{C}^\star$ large enough and*

$$N_1 = \lceil \log n \rceil; \quad \forall 1 \leq i \leq N_1 : \begin{cases} h_n^i = \left(\frac{n}{\sqrt{\log\log n}} \right)^{-2/(4\beta_i+4\sigma+1)} \\ t_{n,i}^2 = \left(\frac{n}{\sqrt{\log\log n}} \right)^{-4\beta_i/(4\beta_i+4\sigma+1)} \end{cases},$$

$$N_2 = \lceil \log\log n/(\bar{r}-\underline{r}) \rceil; \forall 1 \leq i \leq N_2 : \begin{cases} h_n^{N_1+i} = \left(\frac{\log n}{2c} \right)^{-1/r_i}, c < \underline{\alpha} \exp\left(-\frac{1}{\underline{r}} \right) \\ t_{n,N_1+i}^2 = \frac{(\log n)^{(4\sigma+1)/(2r_i)}}{n} \sqrt{\log\log\log n} \end{cases},$$

satisfies (5)*, with testing rate $\Psi_n = \{\psi_{n,\tau}\}_{\tau \in \mathcal{T}}$ given by*

$$\psi_{n,\tau} = \left(\frac{n}{\sqrt{\log\log n}} \right)^{-2\beta/(4\beta+4\sigma+1)} 1_{r=0} + \frac{(\log n)^{(4\sigma+1)/(4r)}}{\sqrt{n}} (\log\log\log n)^{1/4} 1_{r \in [\underline{r}, \bar{r}]}.$$

We note that if Assumptions **(T)** and **(P)** hold for f_0 in $\mathcal{F}(\overline{\alpha}, \bar{r}, \beta_0, L)$, the same optimality proof as in Theorem 1 gives us that the loss of the $\log\log n$ to some power factor is optimal over alternatives in $\bigcup_{\alpha \in [\underline{\alpha}, \overline{\alpha}], \beta \in [\underline{\beta}, \bar{\beta}]} \mathcal{F}(\alpha, 0, \beta, L)$. A loss of a $(\log\log\log n)^{1/4}$ factor appears over alternatives of supersmooth densities (less smooth than f_0) with respect to the minimax rate in [3]. We do not prove that this loss is optimal.

3 Exponentially smooth noise in a semiparametric context

In this section, we assume the noise density g to be exponentially smooth and stable (4), for some **unknown** $s \in [\underline{s}; \bar{s}]$ and fixed (known) bounds $0 < \underline{s} < \bar{s} \leq 2$. More precisely, we suppose that the noise has symmetric stable law having Fourier transform

Assumption (S) $\Phi^g(u) = \exp(-|u|^s)$ where $s \in [\underline{s}; \bar{s}]$.

The results of Section 3.1 are valid under the more general assumption (4) with known scale parameter γ, which enables us to select the smoothness parameter among the wider class of not necessarily symmetric stable densities with known scale parameter. Nevertheless, the exact form of Fourier transform Φ^g is needed for deconvolution purposes (see Section 3.2).

For the model to be identifiable, we must assume that f is not too smooth, i.e. its Fourier transform does not decay asymptotically faster than a known polynomial of order β'.

Assumption (A) There exists some known $A > 0$, such that $|\Phi(u)| \geq A|u|^{-\beta'}$ for large enough $|u|$.

The notation $q_{\beta'}$ is used for the function $u \mapsto A|u|^{-\beta'}$. Under assumptions **(S)** and **(A)** the model is identifiable. Indeed, considering the Fourier transforms, we get for all real number u

$$\log|\Phi^p(u)| = \log|\Phi(u)| - |u|^s.$$

Now assume that we have the equality between two Fourier transform for the observations $\Phi_1^p = \Phi_2^p$, where $\Phi_1^p(u) = \Phi_1(u)e^{-|u|^{s_1}}$ and $\Phi_2^p(u) = \Phi_2(u)e^{-|u|^{s_2}}$. Without loss of generality, we may assume $s_1 \leq s_2$. Then we get

$$|u|^{-s_1}\log|\Phi_1(u)| - 1 = |u|^{-s_1}\log|\Phi_2(u)| - |u|^{s_2-s_1}$$

and taking the limit when $|u|$ tends to infinity implies (with assumption **(A)**) that $s_1 = s_2$ and then $\Phi_1 = \Phi_2$ which proves the identifiability of the model.

In this context, $\mathbb{P}_{f,s}$ and $\mathbb{E}_{f,s}$ respectively denote probability and expectation with respect to the model under parameters (f, s).

3.1 Estimation of the self-similarity index s

We first present a selection procedure $\hat{s}_n$ which asymptotically recovers the true value of the smoothness parameter s, with fast rate of convergence. We use a discrete grid $\{s_1, \ldots, s_N\}$, with a number N of points growing to infinity.

The asymptotic behavior of the Fourier transform Φ^p of the observations is used to select the smoothness index s. More precisely, we have for any large enough $|u|$

$$A|u|^{-\beta'}\exp(-|u|^s) \leq |\Phi^p(u)| \leq \exp(-|u|^s),$$

namely, the function $|\Phi^p|$ asymptotically belongs to the *pipe* $[q_{\beta'}(u)e^{-|u|^s}; e^{-|u|^s}]$. Let us now consider a discrete grid $0 < \underline{s} = s_1 < s_2 < \ldots < s_N = \bar{s} \leq 2$ and denote $\Phi^{[k]}(u) = e^{-|u|^{s_k}}$. These families of functions $\{\Phi^{[k]}\}_{1\leq k\leq N}$ and $\{q_{\beta'}\Phi^{[k]}\}_{1\leq k\leq N}$ form an asymptotically decreasing family as there exists some positive real number u_1 such that for all real u with $|u| \geq u_1$, we have

$$\Phi^{[1]}(u) \geq q_{\beta'}(u)\Phi^{[1]}(u) \geq \Phi^{[2]}(u) \geq \cdots \geq \Phi^{[N]}(u) \geq q_{\beta'}(u)\Phi^{[N]}(u). \tag{10}$$

If the size of the grid is sufficiently small, the modulus of the Fourier transform Φ^p will asymptotically belong to one of the pipes $[q_{\beta'}\Phi^{[k]}; \Phi^{[k]}]$. Our estimation procedure uses the empirical estimator

$$\hat{\Phi}_n^p(u) = \frac{1}{n}\sum_{j=1}^{n} \exp(-iuY_j), \quad \forall u \in \mathbb{R},$$

of the Fourier transform Φ^p at some point u_n which tends to infinity with n. The procedure selects the smoothness parameter among $\{s_1, \ldots, s_N\}$ by choosing the pipe $[q_{\beta'}(u_n)\Phi^{[k]}(u_n); \Phi^{[k]}(u_n)]$ closest to the function $|\hat{\Phi}_n^p(u_n)|$. More precisely

$$\hat{s}_n = \begin{cases} s_k \ \text{ if } \frac{1}{2}\left\{q_{\beta'}\Phi^{[k]} + \Phi^{[k+1]}\right\}(u_n) \leq |\hat{\Phi}_n^p(u_n)| < \frac{1}{2}\left\{q_{\beta'}\Phi^{[k-1]} + \Phi^{[k]}\right\}(u_n) \\ \qquad\quad \text{and } 2 \leq k \leq N-1, \\ s_1 \ \text{ if } |\hat{\Phi}_n^p(u_n)| \geq \frac{1}{2}\left\{q_{\beta'}\Phi^{[1]} + \Phi^{[2]}\right\}(u_n), \\ s_N \ \text{ if } |\hat{\Phi}_n^p(u_n)| < \frac{1}{2}\left\{q_{\beta'}\Phi^{[N-1]} + \Phi^{[N]}\right\}(u_n), \end{cases} \tag{11}$$

where $\{u_n\}_{n\geq 0}$ is a sequence of positive real numbers growing to infinity and to be chosen later. See Figure 1 for an illustration of this procedure.

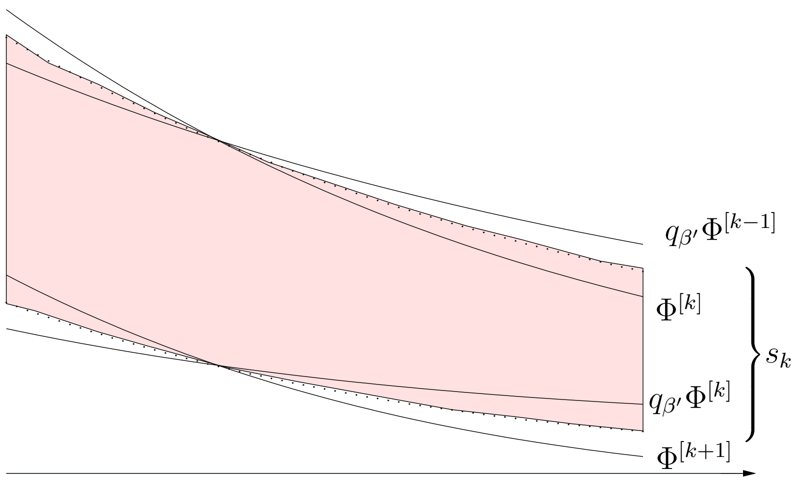

Figure 1. Estimation procedure for s. When $|\hat{\Phi}_n^p(u_n)|$ lies in the grey region, we choose $\hat{s}_n = s_k$.

This estimation procedure is well-defined for large enough n as for any $2 \leq k \leq N-1$, we have $\{q_{\beta'}\Phi^{[k]} + \Phi^{[k+1]}\}(u_n) \leq \{q_{\beta'}\Phi^{[k-1]} + \Phi^{[k]}\}(u_n)$

This procedure is proved to be consistent, with an exponential rate of convergence, in the following proposition.

Proposition 1 *Under assumptions* **(S)** *and* **(A)**, *consider the estimation procedure given by* (11) *where*

$$u_n = \left(\frac{\log n}{2} - \frac{2\beta' + a\bar{s}}{2\bar{s}} \log\log n \right)^{1/\bar{s}},$$

for some fixed $a > 1$ and the equidistant grid $\underline{s} = s_1 < s_2 < \ldots < s_N = \bar{s}$ is chosen as

$$|s_{k+1} - s_k| = d_n = \bar{s}(\log n)^{-1}(\log\log n)^{-1} \quad ; N - 1 = (\bar{s} - \underline{s})/d_n.$$

Then, $\hat{s}_n$ is strongly consistent, i.e.

$$\lim_{n\to\infty} \hat{s}_n = s \ ; \ \mathbb{P}_{f,s} - \textit{almost surely.}$$

Moreover, for each number of observations n, denote by $s_n(s)$ the unique point s_k on the grid such that $s_k \leq s < s_{k+1}$. We have

$$\mathbb{P}_{f,s}(\hat{s}_n \neq s_n(s)) \leq \exp\left(-\frac{A^2}{4}(\log n)^a (1 + o(1)) \right),$$

where A is defined in Assumption **(A)** *and $a > 1$ depends on the choice of u_n.*

Remark 2 *The result remains valid for any sequence d_n satisfying*

$$d_n u_n^{\bar{s}} \log u_n \leq 1 \quad \textit{and} \quad \log(1/d_n) = o((\log n)^a).$$

3.2 Adaptive estimation and tests

For the rest of this section, we shall assume that the unknown density f belongs to some Sobolev class $\mathcal{F}(0, 0, \beta, L)$ where $\beta > 0$ is the smoothness parameter and L is a positive constant. We assume that the unknown parameter β belongs to some known interval $[\underline{\beta}, \bar{\beta}]$.

We now plug the preliminary estimator of s in the usual estimation and testing procedures.

Let us introduce the kernel deconvolution estimator $\hat{K}_n$ built on the preliminary estimation of s and defined by its Fourier transform $\Phi^{\hat{K}_n}$,

$$\Phi^{\hat{K}_n}(u) = \exp\left\{\left(\frac{|u|}{\hat{h}_n}\right)^{\hat{s}_n}\right\} 1_{|u|\leq 1}, \tag{12}$$

$$\text{where } \hat{h}_n = \left(\frac{\log n}{2} - \frac{\bar{\beta} - \hat{s}_n + 1/2}{\hat{s}_n} \log\log n\right)^{-1/\hat{s}_n}. \tag{13}$$

Note that both the bandwidth sequence $\hat{h}_n$ and the kernel $\hat{K}_n$ are random and depend on observations $Y_1, \ldots, Y_n$. Now, the estimator of f is given by

$$\hat{f}_n(x) = \frac{1}{n\hat{h}_n} \sum_{j=1}^{n} \hat{K}_n\left(\frac{Y_j - x}{\hat{h}_n}\right). \tag{14}$$

This estimation procedure is consistent and adaptively achieves the minimax rate of convergence when considering unknown densities f in the union of Sobolev balls $\mathcal{F}(0, 0, \beta, L)$ with $\beta \in [\underline{\beta}, \bar{\beta}] \subset (1/2; +\infty)$ and unknown smoothness parameter $s \in [\underline{s}; \bar{s}]$.

Note that when a function belongs to $\mathcal{F}(0, 0, \beta, L)$ and assumption **(A)** is fulfilled, we necessarily have $\beta' > \beta + 1/2$.

Corollary 1 *Under assumptions* **(S)** *and* **(A)**, *for any $\bar{\beta} > \underline{\beta} > 1/2$, the estimation procedure given by* (14) *which uses estimator $\hat{s}_n$ defined by* (11) *with parameter values: u_n given by Proposition 1 with $a > \bar{s}/\underline{s}$,*

$$d_n = \min\left\{(\log n)^{-(\bar{\beta}-1/2)/\underline{s}}, \bar{s}(\log n \log\log n)^{-1}\right\},$$

satisfies, for any real number x,

$$\limsup_{n\to\infty} \sup_{s\in[\underline{s};\bar{s}]} \sup_{\beta\in[\underline{\beta},\bar{\beta}]} \sup_{f\in\mathcal{F}(0,0,\beta,L)} (\log n)^{(2\beta-1)/s} \mathbb{E}_{f,s}|\hat{f}_n(x) - f(x)|^2 < \infty.$$

Moreover, this rate of convergence is asymptotically adaptive optimal.

Remark 3 *This result is obtained by using that, with high probability, the estimator $\hat{s}_n$ is equal to the point s_k on the grid such that $s_k \leq s < s_{k+1}$ (see Proposition 1). Then, using the deconvolution kernel built on s_k is as good as using the true value s, as soon as the difference $|s_k - s|$ is sufficiently small (which is ensured by the size of the grid). Note that the fact that we underestimate s by using $s_k \leq s$ is rather important as deconvolution with overestimated s would lead to unbounded risk.*

Note that the optimality of this procedure is a direct consequence of a result by [8] where he considers the convolution model for circular data with β and s fixed and known. Therefore we may say that there is no loss due to adaptation neither with respect to s or β.

Using the same kernel estimator (12) and the same random bandwidth (13), we define

$$\hat{T}_n = \frac{2}{n(n-1)} \sum\sum_{1\le k<j\le n} < \frac{1}{\hat{h}_n}\hat{K}_n\left(\frac{\cdot - Y_k}{\hat{h}_n}\right) \ , \ \frac{1}{\hat{h}_n}\hat{K}_n\left(\frac{\cdot - Y_j}{\hat{h}_n}\right) > . \tag{15}$$

Corollary 2 *Under assumptions* **(S)** *and* **(A)***, for any $\bar{\beta} > \underline{\beta} > 0$, the estimation procedure given by* (15) *which uses estimator $\hat{s}_n$ defined by* (11) *with parameter values: u_n given by Proposition 1 with $a > \bar{s}/\underline{s}$,*

$$d_n = \min\left\{(\log n)^{-2\bar{\beta}/\underline{s}}, \bar{s}(\log n \log\log n)^{-1}\right\},$$

satisfies,

$$\limsup_{n\to\infty} \sup_{s\in[\underline{s};\bar{s}]} \sup_{\beta\in[\underline{\beta},\bar{\beta}]} \sup_{f\in\mathcal{F}(0,0,\beta,L)} (\log n)^{2\beta/s}\left\{\mathbb{E}_{f,s}\left|\hat{T}_n - \int f^2\right|^2\right\}^{1/2} < \infty.$$

Moreover, under additional Assumption **(E)***, this rate of convergence is asymptotically adaptive optimal.*

The rate of convergence of this procedure is the same as in the case of known self-similarity index s and known smoothness parameter β. It is thus asymptotically adaptive optimal according to results obtained by [3].

Let us now define, for any $f_0 \in \mathcal{F}(0,0,\bar{\beta},L)$,

$$\hat{T}^0_n = \frac{2}{n(n-1)} \sum\sum_{1\le k<j\le n} < \frac{1}{\hat{h}_n}\hat{K}_n\left(\frac{\cdot - Y_k}{\hat{h}_n}\right) - f_0 \ , \ \frac{1}{\hat{h}_n}\hat{K}_n\left(\frac{\cdot - Y_j}{\hat{h}_n}\right) - f_0 > . \tag{16}$$

This statistic is used for goodness-of-fit testing of the hypothesis

$$\begin{aligned} &H_0 : f = f_0 \\ \text{versus } &H_1(\mathcal{C},\Psi_n) : f \in \cup_{\beta\in[\underline{\beta},\bar{\beta}]}\{f \in \mathcal{F}(0,0,\beta,L) \text{ and } \psi_{n,\beta}^{-2}\|f-f_0\|_2^2 \ge \mathcal{C}\}. \end{aligned}$$

The test is constructed as usual

$$\Delta^\star_n = \begin{cases} 1 \text{ if } |\hat{T}^0_n|\hat{t}_n^{-2} > \mathcal{C}^\star \\ 0 \text{ otherwise,} \end{cases} \tag{17}$$

for some constant $\mathcal{C}^\star > 0$ and a **random** threshold $\hat{t}_n^2$ to be specified.

Corollary 3 *Under assumptions* **(S)** *and* **(A)***, for any $0 < \underline{\beta} < \bar{\beta}$, any $L > 0$ and for any $f_0 \in \mathcal{F}(0,0,\bar{\beta},L)$, consider the testing procedure given by* (17) *which uses the test statistic* (16) *with estimator* $\hat{s}_n$ *defined by* (11) *with parameter values:* u_n *given by Proposition 1 with* $a > 1$,

$$d_n = \min\left\{(\log n)^{-\bar{\beta}/\underline{s}}, \bar{s}(\log n \log\log n)^{-1}\right\},$$

with random threshold and (slightly modified) random bandwidth

$$\hat{t}_n^2 = \left(\frac{\log n}{2}\right)^{-2\bar{\beta}/\hat{s}_n} \quad ; \quad \hat{h}_n = \left(\frac{\log n}{2} - \frac{2\bar{\beta}}{\hat{s}_n}\log\log n\right)^{-1/\hat{s}_n}$$

and any large enough positive constant $\mathcal{C}^\star$. *This testing procedure satisfies* (5) *for any* $\epsilon \in (0, 1)$ *with testing rate*

$$\Psi_n = \{\psi_{n,\beta}\}_{\beta\in[\underline{\beta},\bar{\beta}]} \text{ given by } \psi_{n,\beta} = \left(\frac{\log n}{2}\right)^{-\beta/s}.$$

Moreover, if $f_0 \in \mathcal{F}(0, 0, \bar{\beta}, cL)$ *for some* $0 < c < 1$ *and if Assumptions* **(T)** *and* **(E)** *hold, then this testing rate is asymptotically adaptive optimal over the family of classes* $\{\mathcal{F}(0, 0, \beta, L), \beta \in [\underline{\beta}; \bar{\beta}]\}$ *and for any* $s \in [\underline{s}; \bar{s}]$ *(i.e.* (6) *holds).*

Adaptive optimality (namely (6)) of this testing procedure directly follows from [3] as there is no loss due to adaptation to β nor to s. Note also that the case of known s and adaptation only with respect to β is included in our results and is entirely new.

4 Auxiliary result: Berry-Esseen inequality for degenerate U-statistics of order 2

This section is dedicated to the statement of a non-uniform Berry-Esseen type theorem for degenerate U-statistics. It draws its inspiration from [13] which provides a central limit theorem for degenerate U-statistics. Given a sample $Y_1, \ldots, Y_n$ of i.i.d. random variables, we shall consider U-statistics of the form

$$U_n = \sum\sum_{1\le i<j\le n} H(Y_i, Y_j),$$

where H is a symmetric function. We may assume, without loss of generality, that $\mathbb{E}\{H(Y_1, Y_2)\} = 0$ and thus U_n is centered. We shall focus on **degenerate** U-statistics, namely

$$\mathbb{E}\{H(Y_1, Y_2)|Y_1\} = 0 \text{ , almost surely.}$$

Limit theorems for degenerate U-statistics when H is fixed (independent of the sample size n) are well-known and can be found in any monograph on the subject (see for instance [17]). In that case, the limit distribution is a linear combination of independent and centered $\chi^2(1)$ (chi-square with one degree of freedom) distributions. However, as noticed in [13], a normal distribution may result in some cases where H depends on n. In such a context, [13] provides a central limit theorem. But this result is not enough for our purpose (namely, optimality in Theorem 1). Indeed, we need to control the convergence to zero of the difference between the cumulative distribution function (cdf) of our U-statistic, and the cdf of the Gaussian distribution. Such a result may be derived using classical Martingale methods.

In the rest of this section, n is fixed. Denote by $\mathcal{F}_i$ the σ-field generated by the random variables $\{Y_1, \ldots, Y_i\}$. Define

$$v_n^2 = \mathbb{E}(U_n^2) \quad ; \quad Z_i = \frac{1}{v_n} \sum_{j=1}^{i-1} H(Y_i, Y_j), \quad 2 \le i \le n$$

and note that as the U-statistic is degenerate, we have $\mathbb{E}(Z_i | Y_1, \ldots, Y_{i-1}) = 0$. Thus,

$$S_k = \sum_{i=2}^{k} Z_i, \quad 2 \le k \le n,$$

is a centered Martingale (with respect to the filtration $\{\mathcal{F}_k\}_{k \ge 2}$) and $S_n = v_n^{-1} U_n$. We use a non-uniform Berry-Esseen type theorem for Martingales provided by [14], Theorem 3.9. Denote by ϕ the cdf of the standard Normal distribution and introduce the conditional variance of the increments Z_j's,

$$V_n^2 = \sum_{i=2}^{n} \mathbb{E}(Z_i^2 | \mathcal{F}_{i-1}) = \frac{1}{v_n^2} \sum_{i=2}^{n} \mathbb{E} \left\{ \left(\sum_{j=1}^{i-1} H(Y_i, Y_j) \right)^2 \Bigg| \mathcal{F}_{i-1} \right\}.$$

Theorem 3 *Fix $0 < \delta \le 1$ and define*

$$L_n = \sum_{i=2}^{n} \mathbb{E}|Z_i|^{2+2\delta} + \mathbb{E}|V_n^2 - 1|^{1+\delta}.$$

There exists a positive constant C (depending only on δ) such that for any $0 < \epsilon < 1/2$ and any real x

$$|\mathbb{P}(U_n \le x) - \phi(x/v_n)| \le 16 \epsilon^{1/2} \exp\left(-\frac{x^2}{4 v_n^2} \right) + \frac{C}{\epsilon^{1+\delta}} L_n.$$

5 Proofs

We use C to denote an absolute constant which values may change along the lines.

Proof of Theorem 1 (Upper bound). Let us give the sketch of proof concerning the upper-bound of the test. The statistic T_{n,h^i} will be abbreviated by $T_{n,i}$. We first need to control the first-type error of the test.

$$\begin{aligned}\mathbb{P}_0(\Delta_n^\star = 1) &= \mathbb{P}_0(\exists i \in \{1, \dots, N+1\} \text{ such that } |T_{n,i}| > \mathcal{C}^\star t_{n,i}^2) \\ &\leq \sum_{i=1}^{N+1} \mathbb{P}_0(|T_{n,i} - \mathbb{E}_0(T_{n,i})| > \mathcal{C}^\star t_{n,i}^2 - \mathbb{E}_0(T_{n,i})).\end{aligned}$$

The proof relies on the two following lemmas.

Lemma 1 *For any large enough $\mathcal{C}^\star > 0$, we have*

$$\sum_{i=1}^{N} \mathbb{P}_0(|T_{n,i} - \mathbb{E}_0(T_{n,i})| > \mathcal{C}^\star t_{n,i}^2 - \mathbb{E}_0(T_{n,i})) = o(1).$$

Lemma 2 *For large enough $\mathcal{C}^\star$, there is some $\epsilon \in (0,1)$, such that*

$$\mathbb{P}_0(|T_{n,N+1} - \mathbb{E}_0(T_{n,N+1})| > \mathcal{C}^\star t_{n,N+1}^2 - \mathbb{E}_0(T_{n,N+1})) \leq \epsilon.$$

Lemma 1 relies on the Berry-Esseen type theorem (Theorem 3) presented in Section 4. Its proof is postponed to the very end of the present proof. Proof of Lemma 2 is easy and omitted. *Note for the referee: omitted proofs appear in the appendix.*

Thus, the first type error term is as small as we need, as soon as we choose a large enough constant $\mathcal{C}^\star > 0$ in (9). We now focus on the second-type error of the test. We write

$$\begin{aligned}&\sup_{\tau \in \mathcal{T}} \sup_{f \in \mathcal{F}(\tau,L)} \mathbb{P}_f(\Delta_n^\star = 0) \\ &\quad \leq 1_{\underline{r}>0} \sup_{r \in [\underline{r};\bar{r}], \alpha \geq \underline{\alpha}, \beta \in [\underline{\beta}, \bar{\beta}]} \sup_{\substack{f \in \mathcal{F}(\tau,L) \\ \|f-f_0\|_2^2 \geq \mathcal{C}\psi_{n,\tau}^2}} \mathbb{P}_f(|T_{n,N+1}| \leq \mathcal{C}^\star t_{n,N+1}^2) \\ &\quad + 1_{\underline{r}=\bar{r}=0} \sup_{\alpha \geq \underline{\alpha}, \beta \in [\underline{\beta};\bar{\beta}]} \sup_{\substack{f \in \mathcal{F}(\alpha,0,\beta,L) \\ \|f-f_0\|_2^2 \geq \mathcal{C}\psi_{n,(\alpha,0,\beta)}^2}} \mathbb{P}_f(\forall 1 \leq i \leq N,\ |T_{n,i}| \leq \mathcal{C}^\star t_{n,i}^2).\end{aligned}$$

Note that when the function f in the alternative is supersmooth ($\underline{r} > 0$), we only need the last test (with index $N+1$), whereas when it is ordinary smooth ($\underline{r} = \bar{r} = 0$), we use the family of tests with indexes $i \leq N$. In this second case, we use in fact only the test based on parameter β_f defined as the smallest point on the grid larger than β (see the proof of Lemma 3 below).

Lemma 3 *We have*

$$\sup_{\alpha \geq \underline{\alpha}} \sup_{\beta \in [\underline{\beta};\bar{\beta}]} \sup_{\substack{f \in \mathcal{F}(\alpha,0,\beta,L) \\ \|f-f_0\|_2^2 \geq \mathcal{C}\psi^2_{n,(\alpha,0,\beta)}}} \mathbb{P}_f(\forall 1 \leq i \leq N,\ |T_{n,i}| \leq \mathcal{C}^\star t^2_{n,i}) = o(1).$$

Lemma 4 *Fix $\underline{r} > 0$, for any $\alpha \geq \underline{\alpha}, r \in [\underline{r}; \bar{r}], \beta \in [\underline{\beta}; \bar{\beta}]$. For any $\epsilon \in (0;1)$, there exists some large enough $\mathcal{C}^0$ such that for any $\mathcal{C} > \mathcal{C}^0$ and any $f \in \mathcal{F}(\alpha, r, \beta, L)$ such that $\|f - f_0\|_2^2 \geq \mathcal{C}\psi_{n,(\alpha,r,\beta)}$, we have*

$$\mathbb{P}_f(|T_{n,N+1}| \leq \mathcal{C}^\star t^2_{n,N+1}) \leq \epsilon.$$

The proof of Lemma 3 (resp. 4) is postponed (resp. omitted) to the very end of the present proof. Thus, the second type error of the test converges to zero. This ends the proof of (5). ■

We now present the proofs of the lemmas.

Proof of Lemma 1. Let us set $\rho_n = (\log\log n)^{-1/2}$ and fix $1 \leq i \leq N$. We use the obvious notation $p_0 = f_0 * g$. As we have

$$\begin{aligned} &\mathbb{E}_0(T_{n,i}) = \|K_{h^i} * p_0 - f_0\|_2^2 = \|J_{h^i} * f_0 - f_0\|_2^2, \\ &\text{and } < K_h(\cdot - Y_1) - J_h * f_0, J_h * f_0 - f_0 >= 0 \end{aligned}$$

we easily get

$$T_{n,i} - \mathbb{E}_0(T_{n,i}) = \frac{2}{n(n-1)} \sum\sum_{1 \leq k < j \leq n} < K_{h^i}(\cdot - Y_k) - J_{h^i} * f_0, K_{h^i}(\cdot - Y_j) - J_{h^i} * f_0 > .$$

Let us set

$$H(Y_j, Y_k) = 2\{n(n-1)\}^{-1} < K_{h^i}(\cdot - Y_k) - J_{h^i} * f_0, K_{h^i}(\cdot - Y_j) - J_{h^i} * f_0 >$$

and note that H is a symmetric function with $\mathbb{E}_0\{H(Y_1, Y_2)\} = 0$ and $\mathbb{E}_0\{H(Y_1, Y_2)|Y_1\} = 0$. As a consequence, $T_{n,i} - \mathbb{E}_0(T_{n,i})$ is a degenerate U-statistic. Using Theorem 3 (and the notation of Section 4) to control its cdf, we get that for any $0 < \delta \leq 1$, for any $0 < \varepsilon < 1/2$ and any x

$$\begin{aligned} &|\mathbb{P}_0(T_{n,i} - \mathbb{E}_0(T_{n,i}) > x) - (1 - \phi(x/v_n))| \\ &\qquad \leq 16\varepsilon^{1/2} \exp\left(-\frac{x^2}{4v_n^2}\right) + \frac{C}{\varepsilon^{1+\delta}} \left\{ \sum_{i=2}^n \mathbb{E}_0|Z_i|^{2+2\delta} + \mathbb{E}_0|V_n^2 - 1|^{1+\delta} \right\}, \end{aligned}$$

where $v_n^2 = \mathrm{Var}_0(T_{n,i})$ and

$$Z_i = \frac{1}{v_n} \sum_{j=1}^{i-1} H(Y_i, Y_j) \quad \text{and } V_n^2 = \sum_{i=2}^n \mathbb{E}_0(Z_i^2 | \mathcal{F}_{i-1})$$

as in Section 4. Choose $\delta = 1$ and consider ε as a constant (optimization in ε is not necessary in our context), thus

$$|\mathbb{P}_0(T_{n,i} - \mathbb{E}_0(T_{n,i}) > x) - (1 - \phi(x/v_n))| \\ \leq C \exp\left(-\frac{x^2}{4v_n^2}\right) + C\left\{\sum_{i=2}^{n} \mathbb{E}_0|Z_i|^4 + \mathbb{E}_0|V_n^2 - 1|^2\right\}. \quad (18)$$

We want to apply this inequality at the point $x = \mathcal{C}^\star t_{n,i}^2 - \mathbb{E}_0(T_{n,i})$. First, note that

$$\mathbb{E}_0(T_{n,i}) = \|J_{h^i} * f_0 - f_0\|_2^2 = \frac{1}{2\pi}\int_{|u|>1/(h^i)} |\Phi_0(u)|^2 du \leq L(h^i)^{2\bar{\beta}} \leq L t_{n,i}^2,$$

leading to

$$x \geq (\mathcal{C}^\star - L)t_{n,i}^2 = (\mathcal{C}^\star - L)(n\rho_n)^{-4\beta_i/(4\beta_i+4\sigma+1)}$$

and we choose $\mathcal{C}^\star > L$. Now, the variance term v_n^2 satisfies (see [3])

$$v_n^2 = \mathbb{E}_0(T_{n,i} - \mathbb{E}_0(T_{n,i}))^2 = \frac{C}{n^2(h^i)^{4\sigma+1}}(1 + o(1)).$$

Using the choice of the bandwidth h^i, we obtain a bound of the first term in (18)

$$C\exp\left(-\frac{x^2}{4v_n^2}\right) \leq C\exp\left(-\frac{(\mathcal{C}^\star)^2}{C'}\rho_n^{-2}\right) = C(\log n)^{-b},$$

where $b = (\mathcal{C}^\star)^2/(C')$ can be chosen as large as we need. Let us deal with the other terms appearing in (18). For large enough n,

$$| < K_{h^i}(\cdot - Y_k) - J_{h^i} * f_0, K_{h^i}(\cdot - Y_j) - J_{h^i} * f_0 > | \\ \leq \frac{2}{\pi}\int_{|u|\leq 1/h^i} |\Phi^g(u)|^{-2} du \leq \frac{C}{(h^i)^{2\sigma+1}}$$

and thus, for any $p \geq 2$,

$$\mathbb{E}_0\{|H(Y_1, Y_2)|^{2p}\} \leq Cn^{-4p}(h^i)^{-2p(2\sigma+1)}.$$

This leads to

$$\sum_{i=2}^{n} \mathbb{E}_0|Z_i|^4 \leq \frac{1}{v_n^4}\sum_{i=2}^{n}\left(\sum_{j=1}^{i-1}\mathbb{E}_0(H(Y_i,Y_j)^4) + 3\sum\sum_{1\leq j\neq k\leq i-1}\mathbb{E}_0(H(Y_i,Y_j)^2 H(Y_i,Y_k)^2)\right)$$
$$\leq \frac{1}{v_n^4}\sum_{i=2}^{n}\Big((i-1)\mathbb{E}_0(H(Y_1,Y_2)^4) + 3(i-1)(i-2)\mathbb{E}_0(H(Y_1,Y_2)^2 H(Y_1,Y_3)^2)\Big)$$
$$\leq \frac{O(1)}{v_n^4}n^2\mathbb{E}_0(H(Y_1,Y_2)^4) + \frac{O(1)}{v_n^4}n^3\mathbb{E}_0(H(Y_1,Y_2)^2 H(Y_1,Y_3)^2)$$
$$\leq O(1)\frac{n^3}{n^8(h^i)^{4(2\sigma+1)}}n^4(h^i)^{2(4\sigma+1)} = \frac{O(1)}{n(h^i)^2}.$$

Moreover, following the lines of the proof of Theorem 1 in [13] we get

$$\mathbb{E}_0|V_n^2 - 1|^2 \leq \frac{1}{v_n^4}\left(\mathbb{E}_0(G^2(Y_1, Y_2)) + \frac{1}{n}\mathbb{E}_0(H^4(Y_1, Y_2))\right),$$

where $G(x, y) = \mathbb{E}_0(H(Y_1, x)H(Y_1, y))$. In [1] this last term was bounded from above for this model by Ch^i so

$$\mathbb{E}_0|V_n^2 - 1|^2 \leq Ch^i.$$

Returning to (18) we finally get for $x = \mathcal{C}^\star t_{n,i}^2 - \mathbb{E}_0(T_{n,i})$,

$$|\mathbb{P}_0(T_{n,i} - \mathbb{E}_0(T_{n,i}) > x) - \{1 - \phi(x/v_n)\}| \leq C\left((\log n)^{-b} + h^i\right) \leq C(\log n)^{-b}.$$

Finally we obtain, for b large when $\mathcal{C}^\star$ is large

$$\sum_{i=1}^N \mathbb{P}_0(|T_{n,i} - \mathbb{E}_0(T_{n,i})| > \mathcal{C}^\star t_{n,i}^2 - \mathbb{E}_0(T_{n,i})) \leq N(1 - \phi(x/v_n) + C(\log n)^{-b})$$
$$\leq CN\left(v_n x^{-1}\exp(-x^2/(2v_n^2)) + (\log n)^{-b}\right) \leq CN\rho_n(\log n)^{-b} \leq C\frac{(\log\log n)^{-1/2}}{\log n^{b-1}}.$$

■

Proof of Lemma 3. When $\bar{r} = \underline{r} = 0$, let us fix some constant $\mathcal{C} > \mathcal{C}^0$ ($\mathcal{C}^0$ will be chosen later) and a density f belonging to $\mathcal{F}(\alpha, 0, \beta, L)$ for some unknown $\alpha > \underline{\alpha}$ and $\beta \in [\underline{\beta}; \bar{\beta}]$ which satisfies $\|f - f_0\|_2^2 \geq \mathcal{C}\psi_{n,(\alpha,0,\beta)}^2$ (choose β as the largest one). In this proof, we abbreviate $\psi_{n,(\alpha,0,\beta)}$ to $\psi_{n,\beta}$ since in this case, the rate only depends on β. We define β_f as the smallest point on the finite grid $\{\underline{\beta} = \beta_1 < \beta_2 < \ldots < \beta_N = \bar{\beta}\}$ such that $\beta \leq \beta_f$

$$\beta_f \in \{\underline{\beta} = \beta_0 < \beta_1 < \ldots < \beta_N = \bar{\beta}\},\ \ f \in \mathcal{F}(\alpha, 0, \beta, L), \|f - f_0\|_2^2 \geq \mathcal{C}\psi_{n,\beta}^2,$$
$$\beta \leq \beta_f \text{ and } \forall \beta_i < \beta_f, \text{ we have } \beta > \beta_i. \quad (19)$$

We shall abbreviate to h_f, $t_{n,f}^2$ and $T_{n,f}$ the bandwidth, the threshold (both defined in Theorem 1) and the statistic (8) corresponding to parameter β_f. We write

$$\begin{aligned}
&\mathbb{P}_f(\forall i \in \{1, \ldots, N\},\ \ |T_{n,i}| \leq \mathcal{C}^\star t_{n,i}^2)\\
\leq &\mathbb{P}_f(|T_{n,f} - \mathbb{E}_f(T_{n,f})| \geq -\mathcal{C}^\star t_{n,f}^2 + \mathbb{E}_f(T_{n,f}))\\
\leq &\mathbb{P}_f(|T_{n,f} - \mathbb{E}_f(T_{n,f})| \geq \|f - f_0\|_2^2 - \mathcal{C}^\star t_{n,f}^2 + B_f(T_{n,f})), \qquad (20)
\end{aligned}$$

where

$$B_f(T_{n,f}) = \mathbb{E}_f(T_{n,f}) - \|f - f_0\|_2^2 = \|J_h * f\|_2^2 - \|f\|_2^2 + 2\langle f - J_h * f, f_0\rangle$$

is in fact a bias term. It satisfies

$$
\begin{aligned}
|B_f(T_{n,f})| &\leq \int_{|u|\geq 1/h_f} |\Phi(u)|^2 du + 2\Big(\int_{|u|\geq 1/h_f} |\Phi(u)|^2 du \int_{|u|\geq 1/h_f} |\Phi_0(u)|^2 du\Big)^{1/2} \\
&\leq L e^{-2\underline{\alpha}}(h_f^{2\beta} + 2h_f^{\bar{\beta}+\beta}) \leq 3e^{-2\underline{\alpha}} L h_f^{2\beta},
\end{aligned}
$$

as f belongs to $\mathcal{F}(\alpha, 0, \beta, L) \subseteq \mathcal{F}(\underline{\alpha}, 0, \beta, L)$.

Let us study the variance term $\mathbb{E}_f(T_{n,f} - \mathbb{E}_f(T_{n,f}))^2$. According to [3], this term is upper-bounded by $w_{n,f}^2$ given by

$$
\mathbb{E}_f(T_{n,f} - \mathbb{E}_f(T_{n,f}))^2 \leq \frac{C}{n^2 h_f^{4\sigma+1}} + \frac{4\Omega_g^2(f-f_0)}{n} 1_{\beta\geq\sigma} = w_{n,f}^2,
$$

and $\Omega_g(f-f_0)$ is a constant depending on f and g (but not n) and satisfying $|\Omega_g^2(f-f_0)| \leq C\|f-f_0\|_2^{2-2\sigma/\beta}$ (see proof of Theorem 6 in [3]).

Using Markov's inequality, this leads to the following upper bound of (20)

$$
\frac{w_{n,f}^2}{(\|f-f_0\|_2^2 - \mathcal{C}^\star t_{n,f}^2 - 3e^{-2\underline{\alpha}} L h_f^{2\beta})^2}.
$$

We will proceed differently when $\beta < \sigma$ and when $\beta \geq \sigma$. Let us first consider the term concerning $\beta < \sigma$. The point is to use that f satisfies $\|f-f_0\|_2^2 \geq \mathcal{C}\psi_{n,\beta}^2$. Note that we have $\beta_f \geq \beta$, constants $\mathcal{C} > \mathcal{C}^\star$ and

$$
\psi_{n,\beta}^2 t_{n,f}^{-2} = (n\rho_n)^{4(\beta_f-\beta)(4\sigma+1)/\{(4\beta_f+4\sigma+1)(4\beta+4\sigma+1)\}},
$$

ensuring that the term $\mathcal{C}\psi_{n,\beta}^2 - \mathcal{C}^\star t_{n,f}^2$ is always positive. Moreover, as $0 \geq \beta - \beta_f \geq -(\bar{\beta} - \underline{\beta})/\log n$, we have

$$
\begin{aligned}
\psi_{n,\beta}^2 h_f^{-2\beta} &= \exp\left\{ \frac{16\beta(\beta-\beta_f)}{(4\beta_f+4\sigma+1)(4\beta+4\sigma+1)} \log(n\rho_n) \right\} \\
&\geq \exp\left\{ -\frac{16\bar{\beta}(\bar{\beta}-\underline{\beta})}{(4\underline{\beta}+4\sigma+1)^2}(1+o(1)) \right\} =: \mathcal{C}_1(1+o(1)).
\end{aligned}
$$

Thus, we choose $\mathcal{C}^0 = \mathcal{C}^\star + 3e^{-2\underline{\alpha}} L/\mathcal{C}_1$ such that for any $\mathcal{C} > \mathcal{C}^0$, we have

$$
\|f-f_0\|_2^2 - \mathcal{C}^\star t_{n,f}^2 - 3e^{-2\underline{\alpha}} L h_f^{2\beta} \geq (\mathcal{C} - \mathcal{C}^* - 3e^{-2\underline{\alpha}} L/\mathcal{C}_1)\psi_{n,\beta}^2 = a\psi_{n,\beta}^2,
$$

with $a > 0$. Thus, we get

$$\sup_{\alpha > \underline{\alpha}} \sup_{\beta \in [\underline{\beta};\bar{\beta}]} \sup_{\substack{f \in \mathcal{F}(\alpha,0,\beta,L) \\ \|f-f_0\|_2^2 \geq \mathcal{C}\psi_{n,\beta}^2}} \mathbb{P}_f(\forall i \in \{1,\ldots,N\}, \quad |T_{n,i}| \leq \mathcal{C}^\star t_{n,i}^2)$$
$$\leq \max\left\{ \sup_{\beta<\sigma} \sup_f \frac{C}{n^2 h_f^{4\sigma+1} \psi_{n,\beta}^4} ; \sup_{\beta\geq\sigma} \sup_f \frac{C\|f-f_0\|_2^{2-2\sigma/\beta}}{n(\|f-f_0\|_2^2 - \mathcal{C}^\star t_{n,f}^2 - 3e^{-2\underline{\alpha}} L h_f^{2\beta})^2} \right\}.$$

Finally, this leads to the bound

$$\max\left\{ \sup_{\beta<\sigma} \sup_f \frac{C}{n^2 h_f^{4\sigma+1} \psi_{n,\beta}^4} ; \sup_{\beta\geq\sigma} \frac{C}{n\|f-f_0\|_2^{2+2\sigma/\beta} (a/\mathcal{C})^2} \right\}$$
$$\leq \max\left\{ \sup_{\beta<\sigma} \sup_f \frac{C}{n^2 h_f^{4\sigma+1} \psi_{n,\beta}^4} ; \sup_{\beta\geq\sigma} \frac{C}{n\psi_{n,\beta}^{2+2\sigma/\beta}} \right\} \leq \rho_n,$$

which converges to zero as n tends to infinity. ■

Proof of Theorem 1 (Lower bound).

As we already noted after the theorem statement, our test procedure attains the minimax rate associated to the class $\mathcal{F}(\alpha_0, 0, \bar{\beta}, L)$ where f_0 belongs, whenever the alternative f belongs to classes of functions smoother than f_0. Therefore, the lower bound we need to prove concerns the optimality of the loss of order $(\log\log n)^{1/2}$ due to alternatives less smooth than f_0.

More precisely, we prove (6), where the alternative $H_1(\mathcal{C}, \Psi_n)$ is now restricted to $\cup_{\beta\in[\underline{\beta},\bar{\beta}]}\{f \in \mathcal{F}(0,0,\beta,L)$ and $\psi_{n,\beta}^{-2}\|f-f_0\|_2^2 \geq \mathcal{C}\}$ and $\psi_{n,\beta}$ denotes the rate $\psi_{n,\tau}$ when $\tau = (0,0,\beta,L)$.

The general approach for proving such a lower bound (6) is to exhibit a finite number of regularities $\{\beta_k\}_{1\leq k\leq K}$ and corresponding probability distributions $\{\pi_k\}_{1\leq k\leq K}$ on the alternatives $H_1(\mathcal{C}, \psi_{n,\beta_k})$ (more exactly, on parametric subsets of these alternatives) such that the distance between the distributions induced by f_0 (the density being tested) and the mean distribution of the alternatives is small.

We use a finite grid $\bar{\mathcal{B}} = \{\beta_1 < \beta_2 < \ldots < \beta_K\} \subset [\underline{\beta}, \bar{\beta}]$ such that

$$\forall \beta \in [\underline{\beta}, \bar{\beta}], \exists k : |\beta_k - \beta| \leq \frac{1}{\log n}.$$

To each point β in this grid, we associate a bandwidth

$$h_\beta = (n\rho_n)^{-\frac{2}{4\beta+4\sigma+1}}, \rho_n = (\log\log n)^{-1/2}, \quad \text{and} \quad M_\beta = h_\beta^{-1}.$$

We use the same deconvolution kernel as in [3], constructed as follows. Let G be defined as in Lemma 2 in [3]. The function G is an infinitely differentiable function, compactly supported on $[-1,\ 0]$ and such that $\int G = 0$. Then, the deconvolution kernel H_β is defined via its Fourier transform Φ^{H_β} by

$$\Phi^{H_\beta}(u) = \Phi^G(h_\beta u)(\Phi^g(u))^{-1}.$$

Note that the factor ρ_n in the bandwidth's expression corresponds to the loss for adaptation.

We also consider for each β, a probability distribution π_β (also denoted π_k when $\beta = \beta_k$) defined on $\{-1,+1\}^{M_\beta}$ which is in fact the product of Rademacher distributions on $\{-1,+1\}$ and a parametric subset of $H_1(\mathcal{C}, \psi_{n,\beta})$ containing the following functions

$$f_{\theta,\beta}(x) = f_0(x) + \sum_{j=1}^{M_\beta} \theta_j h_\beta^{\beta+\sigma+1} H_\beta\left(x - x_{j,\beta}\right), \qquad \begin{cases} \theta_j \text{ i.i.d. with } \mathbb{P}(\theta_j = \pm 1) = 1/2, \\ x_{j,\beta} = j h_\beta \in [0,1]. \end{cases}$$

Convolution of these functions with g induces another parametric set of functions

$$p_{\theta,\beta}(y) = p_0(y) + \sum_{j=1}^{M_\beta} \theta_j h_\beta^{\beta+\sigma+1} G_\beta\left(y - x_{j,\beta}\right)$$

where $G_\beta(y) = h_\beta^{-1} G\left(y/h_\beta\right) = H_\beta * g(y)$.

As established in [3] (Lemmas 2 and 4), for any β, any $\theta \in \{-1,+1\}^{M_\beta}$ and small enough h_β (i.e. large enough n) the function $f_{\theta,\beta}$ is a probability density and belongs to the Sobolev class $\mathcal{F}(0,0,\beta,L)$ and $p_{\theta,\beta}$ is also a probability density. Moreover we have

$$\frac{1}{K} \sum_{\beta \in \bar{\mathcal{B}}} \pi_\beta \left(\|f_{\theta,\beta} - f_0\|_2^2 \geq \mathcal{C} \psi_{n,\beta}^2 \right) \underset{n \to +\infty}{\longrightarrow} 1,$$

which means that for each β, the random parametric family $\{f_{\theta,\beta}\}_\theta$ belongs almost surely (with respect to the measure π_β) to the alternative set $H_1(\mathcal{C}, \psi_{n,\beta})$. The subset of functions which are not in the alternative $H_1(\mathcal{C}, \Psi_n)$ is asymptotically negligible. We then have,

$$
\begin{aligned}
\gamma_n &\triangleq \inf_{\Delta_n}\left\{\mathbb{P}_0(\Delta_n=1)+\sup_{f\in H_1(\mathcal{C},\Psi_n)}\mathbb{P}_f(\Delta_n=0)\right\}\\
&\geq \inf_{\Delta_n}\left\{\mathbb{P}_0(\Delta_n=1)+\frac{1}{K}\sum_{k=1}^{K}\sup_{f\in H_1(\mathcal{C},\psi_{n,\beta_k})}\mathbb{P}_f(\Delta_n=0)\right\}\\
&\geq \inf_{\Delta_n}\left\{\mathbb{P}_0(\Delta_n=1)+\frac{1}{K}\sum_{k=1}^{K}\Big(\int_\theta \mathbb{P}_{f_{\theta,\beta_k}}(\Delta_n=0)\pi_k(d\theta)\right.\\
&\qquad\qquad\qquad\qquad\qquad\left.-\pi_k(\|f_{\theta,\beta_k}-f_0\|_2^2<\mathcal{C}\psi^2_{n,\beta_k})\Big)\right\}\\
&\geq \inf_{\Delta_n}\left\{\mathbb{P}_0(\Delta_n=1)+\frac{1}{K}\sum_{k=1}^{K}\Big(\int_\theta \mathbb{P}_{f_{\theta,\beta_k}}(\Delta_n=0)\pi_k(d\theta)\Big)\right\}+o(1).
\end{aligned}
$$

Let us denote by

$$
\pi=\frac{1}{K}\sum_{k=1}^{K}\pi_k \quad\text{ and }\quad \mathbb{P}_\pi=\frac{1}{K}\sum_{k=1}^{K}\mathbb{P}_k=\frac{1}{K}\sum_{k=1}^{K}\int_\theta \mathbb{P}_{f_{\theta,\beta_k}}\,\pi_k(d\theta).
$$

Those notations lead to

$$
\begin{aligned}
\gamma_n &\geq \inf_{\Delta_n}\{\mathbb{P}_0(\Delta_n=1)+\mathbb{P}_\pi(\Delta_n=0)\}\\
&\geq \inf_{\Delta_n}\left\{1-\int_{\Delta_n=0}d\mathbb{P}_0+\int_{\Delta_n=0}d\mathbb{P}_\pi\right\}\geq 1-\sup_A\int_A(d\mathbb{P}_0-d\mathbb{P}_\pi)\\
&\geq 1-\frac{1}{2}\|\mathbb{P}_\pi-\mathbb{P}_0\|_1, \qquad (21)
\end{aligned}
$$

where we used Scheffé's Lemma.

The finite grid $\bar{\mathcal{B}}$ is split into subsets $\bar{\mathcal{B}}=\cup_l\bar{\mathcal{B}}_l$ with $\bar{\mathcal{B}}_l\cap\bar{\mathcal{B}}_k=\emptyset$ when $l\neq k$ and such that

$$
\forall l,\ \ \forall\beta_1\neq\beta_2\in\bar{\mathcal{B}}_l,\ \ \frac{c\log\log n}{\log n}\leq|\beta_1-\beta_2|.
$$

The number of subsets $\bar{\mathcal{B}}_l$ is denoted by $K_1=O(\log\log n)$ and the cardinality $|\bar{\mathcal{B}}_l|$ of each subset $\bar{\mathcal{B}}_l$ is of the order $O(\log n/\log\log n)$, uniformly with respect to l.

The lower bound (6) is then obtained from (21) in the following way

$$
\gamma_n\geq 1-\frac{1}{2K_1}\sum_{l=1}^{K_1}\left\|\frac{1}{|\bar{\mathcal{B}}_l|}\sum_{\beta\in\bar{\mathcal{B}}_l}\mathbb{P}_\beta-\mathbb{P}_0\right\|_1,
$$

where $\mathbb{P}_\beta=\int_\theta\mathbb{P}_{f_{\theta,\beta}}\pi_\beta(d\theta)$.

Here we do not want to apply the triangular inequality to the whole set of indexes $\bar{\mathcal{B}}$. Indeed, this would lead to a lower bound equal to 0. Yet, if we do not apply some sort of triangular inequality, we cannot deal with the sum because of too much dependency. This is why we introduced the subsets $\bar{\mathcal{B}}_l$ with the property that two points in the same subset $\bar{\mathcal{B}}_l$ are far enough away from each other. This technique was already used in [12] for the discrete regression model.

Let us denote by ℓ_β the likelihood ratio

$$\ell_\beta = \frac{d\mathbb{P}_\beta}{d\mathbb{P}_0} = \int \frac{d\mathbb{P}_{f_{\theta,\beta}}}{d\mathbb{P}_0}\, \pi_\beta\,(d\theta).$$

We thus have

$$\gamma_n \geq 1 - \frac{1}{2K_1}\sum_{l=1}^{K_1}\int\left(\frac{1}{|\bar{\mathcal{B}}_l|}\sum_{\beta\in\bar{\mathcal{B}}_l}\ell_\beta - 1\right)d\mathbb{P}_0 = 1 - \frac{1}{2K_1}\sum_{l=1}^{K_1}\left\|\frac{1}{|\bar{\mathcal{B}}_l|}\sum_{\beta\in\bar{\mathcal{B}}_l}\ell_\beta - 1\right\|_{\mathbb{L}_1(\mathbb{P}_0)}.$$

Now we use the usual inequality between $\mathbb{L}_1$ and $\mathbb{L}_2$-distances to get that

$$\gamma_n \geq 1 - \frac{1}{2K_1}\sum_{l=1}^{K_1}\left\|\frac{1}{|\bar{\mathcal{B}}_l|}\sum_{\beta\in\bar{\mathcal{B}}_l}\ell_\beta - 1\right\|_{\mathbb{L}_2(\mathbb{P}_0)} = 1 - \frac{1}{2K_1}\sum_{l=1}^{K_1}\left\{\mathbb{E}_0\left(\frac{1}{|\bar{\mathcal{B}}_l|}\sum_{\beta\in\bar{\mathcal{B}}_l}\ell_\beta - 1\right)^2\right\}^{1/2}.$$

Let us focus on the expected value appearing in the lower bound. We have

$$\mathbb{E}_0\left(\frac{1}{|\bar{\mathcal{B}}_l|}\sum_{\beta\in\bar{\mathcal{B}}_l}\ell_\beta - 1\right)^2 = \frac{1}{|\bar{\mathcal{B}}_l|^2}\sum_{\beta\in\bar{\mathcal{B}}_l}Q_\beta + \frac{1}{|\bar{\mathcal{B}}_l|^2}\sum_{\substack{\beta,\nu\in\bar{\mathcal{B}}_l\\ \beta\neq\nu}}Q_{\beta,\nu},$$

where there are two quantities to evaluate

$$Q_\beta = \mathbb{E}_0\left((\ell_\beta - 1)^2\right) \quad \text{and} \quad Q_{\beta,\nu} = \mathbb{E}_0\left(\ell_\beta\ell_\nu - 1\right).$$

The first term Q_β is treated as in [3]. It corresponds to the computation of a χ^2-distance between the two models induced by $\mathbb{P}_\beta$ and $\mathbb{P}_0$ (see term Δ^2 in [3]). Indeed we have

$$Q_\beta \leq CM_\beta n^2 h_\beta^{4\beta+4\sigma+2} \leq C\frac{1}{\rho_n^2}.$$

This upper bound goes to infinity very slowly. The number of β's in each $\bar{\mathcal{B}}_l$ compensates this behaviour

$$\frac{1}{|\bar{\mathcal{B}}_l|^2}\sum_{\beta\in\bar{\mathcal{B}}_l}Q_\beta \leq \frac{1}{|\bar{\mathcal{B}}_l|\rho_n^2} = O\left(\frac{(\log\log n)^2}{\log n}\right) = o(1).$$

The second term is a new one (with respect to non-adaptive case). As G is compactly supported and the points β and ν are far away from each other, we can prove that this term is asymptotically negligible. Recall the expression of the likelihood ratio for a fixed β

$$\ell_\beta = \int \frac{d\mathbb{P}_{f_{\theta,\beta}}}{d\mathbb{P}_0}\,\pi_\beta\left(d\theta\right) = \int \prod_{r=1}^{n}\left(1 + \sum_{j=1}^{M_\beta}\theta_{j,\beta}h_\beta^{\beta+\sigma+1}\frac{G_\beta\left(Y_r - x_{j,\beta}\right)}{p_0\left(Y_r\right)}\right)\pi_\beta\left(d\theta\right).$$

Thus,

$$\begin{aligned}\ell_\beta\ell_\nu &= \int \frac{d\mathbb{P}_{f_{\theta,\beta}}}{d\mathbb{P}_0}\,\pi_\beta\left(d\theta\right)\int \frac{d\mathbb{P}_{f_{\theta,\nu}}}{d\mathbb{P}_0}\,\pi_\nu\left(d\theta\right)\\ &= \int \prod_{r=1}^{n}\left(1 + \sum_{j=1}^{M_\beta}\theta_{j,\beta}\,h_\beta^{\beta+\sigma+1}\frac{G_\beta\left(Y_r - x_{j,\beta}\right)}{p_0\left(Y_r\right)}\right)\\ &\quad\times\left(1 + \sum_{i=1}^{M_\nu}\theta_{i,\nu}h_\nu^{\nu+\sigma+1}\frac{G_\nu\left(Y_r - x_{i,\nu}\right)}{p_0\left(Y_r\right)}\right)\pi_\beta\left(d\theta_{.,\beta}\right)\pi_\nu\left(d\theta_{.,\nu}\right).\end{aligned}$$

The random variables Y_r are i.i.d. and $\mathbb{E}_0\left(\frac{G_\beta\left(Y_r - x_{j,\beta}\right)}{p_0\left(Y_r\right)}\right) = 0$. Thus we have

$$\begin{aligned}\mathbb{E}_0\left(\ell_\beta\ell_\nu\right) = \int \Bigg[1 + \sum_{j=1}^{M_\beta}\sum_{i=1, i\subset j}^{M_\nu}\theta_{j,\beta}\theta_{i,\nu}h_\beta^{\beta+\sigma+1}h_\nu^{\nu+\sigma+1}\\ \mathbb{E}_0\left(\frac{G_\beta\left(Y_1 - x_{j,\beta}\right)G_\nu\left(Y_1 - x_{i,\nu}\right)}{p_0^2\left(Y_1\right)}\right)\Bigg]^n\pi_\beta\left(d\theta_{.,\beta}\right)\pi_\nu\left(d\theta_{.,\nu}\right).\end{aligned}$$

where the second sum concerns only some indexes i, denoted by $i \subset j$. This notation stands for the set of indexes i such that $[(i-1)h_\beta; ih_\beta]\cap[(j-1)h_\nu; jh_\nu] \neq \emptyset$. From now on, we fix $\beta > \nu$. Denote by G' (resp. p_0') the first derivative of G (resp. p_0). (The density p_0 is continuously differentiable as it is the convolution product $f_0 * g$ where the noise density g is at least continuously differentiable).

Lemma 5 *For any $\beta > \nu$ and any $(i,j) \in \{1,\ldots,M_\nu\}\times\{1,\ldots,M_\beta\}$, we have*

$$\mathbb{E}_0\left(\frac{G_\beta\left(Y_1 - x_{j,\beta}\right)G_\nu\left(Y_1 - x_{i,\nu}\right)}{p_0^2\left(Y_1\right)}\right) = \frac{h_\nu}{h_\beta^2}R_{i,j},$$

where R_{ij} satisfies

$$|R_{i,j}| \leq (\inf_{[0,1]} p_0)^{-1}\|G\|_\infty\|G'\|_\infty(1 + o(1))$$

and $o(1)$ is uniform with respect to (i,j).

The proof of this lemma is omitted. Applying Lemma 5, we get

$$Q_{\beta,\nu}+1 = \int \left[1 + \sum_{j=1}^{M_\beta} \sum_{i=1, i\subset j}^{M_\nu} \theta_{j,\beta}\theta_{i,\nu} h_\beta^{\beta+\sigma+1} h_\nu^{\nu+\sigma+1} \frac{h_\nu}{(h_\beta)^2} R_{i,j}\right]^n \pi_\beta\left(d\theta_{.,\beta}\right)\pi_\nu\left(d\theta_{.,\nu}\right).$$

Lemma 6 *Let U be a real valued random variable such that $\forall k \in \mathbb{N},\ \mathbb{E}\left(U^{2k+1}\right) = 0$. We have, for any integer $n \geq 1$,*

$$\mathbb{E}\left(1+U\right)^n \leq 1 + \sum_{k=1}^{\lfloor \frac{n}{2} \rfloor} \frac{n^{2k}}{(2k)!}\mathbb{E}\left(U^{2k}\right),$$

where $\lfloor x \rfloor$ is the largest integer which is smaller than x.

The proof is obvious and therefore omitted. Apply Lemma 6 to get the inequality

$$Q_{\beta,\nu} \leq \sum_{k=1}^{\lfloor \frac{n}{2} \rfloor} \frac{n^{2k}}{(2k)!}(h_\beta^{\beta+\sigma-1} h_\nu^{\nu+\sigma+2})^{2k} \mathbb{E}_\pi \left(\sum_{j=1}^{M_\beta} \sum_{i=1,i\subset j}^{M_\nu} \theta_{j,\beta}\theta_{i,\nu} R_{i,j}\right)^{2k}.$$

But the θ's are i.i.d. Rademacher variables and the $R_{i,j}$'s are deterministic, thus

$$\mathbb{E}_\pi \left(\sum_{j=1}^{M_\beta} \sum_{i=1,i\subset j}^{M_\nu} \theta_{j,\beta}\theta_{i,\nu} R_{i,j}\right)^{2k} = \sum_{1\leq j_1,\ldots,j_k \leq M_\beta} \sum_{\substack{1\leq i_1,\ldots,i_k\leq M_\nu \\ \forall l, i_l \subset j_l}} \left(\prod_{l=1}^{k} R^2_{i_l,j_l}\right).$$

Using the bound on the $R_{i,j}$ given by Lemma 5,

$$\mathbb{E}_\pi \left(\sum_{j=1}^{M_\beta} \sum_{i=1,i\subset j}^{M_\nu} \theta_{j,\beta}\theta_{i,\nu} R_{i,j}\right)^{2k} \leq \left((\inf_{[0,1]} p_0)^{-1} \|G\|_\infty \|G'\|_\infty (1+o(1))\right)^{2k} h_\nu^k.$$

Indeed, each index j_l may take at most $M_\beta = h_\beta^{-1}$ different values but the constraint $i_l \subset j_l$ implies that each index i_l is limited to at most h_β/h_ν different values. Thus we get

$$\begin{aligned} Q_{\beta,\nu} &\leq C \sum_{k=1}^{\lfloor \frac{n}{2} \rfloor} \frac{n^{2k}}{(2k)!} \left(C h_\beta^{\beta+\sigma+1} h_\nu^{\nu+\sigma+1} \frac{h_\nu}{h_\beta^2}\right)^{2k} h_\nu^{-k} \\ &\leq C \sum_{k=1}^{\lfloor \frac{n}{2} \rfloor} \left(n^2 h_\beta^{2\beta+2\sigma+1/2} h_\nu^{2\nu+2\sigma+1/2} \frac{h_\nu^{5/2}}{h_\beta^{5/2}}\right)^k \leq C \sum_{k=1}^{\lfloor \frac{n}{2} \rfloor} \left(\frac{h_\nu^{5/2}}{\rho_n^2 h_\beta^{5/2}}\right)^k \leq C \frac{1}{\rho_n^2} \frac{h_\nu^{5/2}}{h_\beta^{5/2}}. \end{aligned}$$

As $\beta > \nu$ both belong to some set $\bar{\mathcal{B}}_l$, we have $\beta - \nu \geq c(\log\log n)/(\log n)$ and according to the choice of the bandwidths,

$$\frac{h_\nu^{5/2}}{h_\beta^{5/2}} = (n\rho_n)^{-\frac{20(\beta-\nu)}{(4\beta+4\sigma+1)(4\nu+4\sigma+1)}} \leq \exp\left\{-\frac{20\, c \log\log n}{(4\bar{\beta}+4\sigma+1)^2}(1+o(1))\right\} \leq (\log n)^{-w},$$

where the constant w (depending on the constant c used in the construction of the sets $\bar{\mathcal{B}}_l$) can be tailored to our need. Therefore

$$\frac{1}{|\bar{\mathcal{B}}_l|^2} \sum_{\substack{\beta,\nu \in |\bar{\mathcal{B}}_l| \\ \beta \neq \nu}} Q_{\beta,\nu} \leq \frac{C}{\rho_n^2 (\log n)^w}$$

which goes to 0 as n goes to $+\infty$. We finally obtain the upper bound

$$\mathbb{E}_0 \left(\left(\frac{1}{|\bar{\mathcal{B}}_l|} \sum_{\beta \in |\bar{\mathcal{B}}_l|} \ell_\beta - 1 \right)^2 \right) \leq O\left(\frac{1}{|\bar{\mathcal{B}}_l| \rho_n^2} \right) + O\left(\frac{1}{\rho_n^2 (\log n)^w} \right) = o(1),$$

which leads to

$$\gamma_n \geq 1 - \frac{1}{2} \frac{1}{K_1} \sum_{l=1}^{K_1} \left\{ O\left(\frac{1}{|\bar{\mathcal{B}}_l| \rho_n^2} \right) + O\left(\frac{1}{\rho_n^2 (\log n)^c} \right) \right\}^{1/2} = 1 + o(1).$$

■

Proof of Proposition 1. We fix $\epsilon > 0$. Now,

$$\mathbb{P}_{f,s}(|\hat{s}_n - s| \geq \epsilon) \leq \mathbb{P}_{f,s}(\hat{s}_n \neq s_n(s)) + \mathbb{P}_{f,s}(|s - s_n(s)| \geq \epsilon).$$

As $|s - s_n(s)| \leq d_n$ which converges to zero, we get that for large enough n, the term $\mathbb{P}_{f,s}(|s - s_n(s)| \geq \epsilon)$ is equal to zero. Let us now consider the term $\mathbb{P}_{f,s}(\hat{s}_n \neq s_n(s)) = \mathbb{P}_{f,s}(\hat{s}_n > s_n(s)) + \mathbb{P}_{f,s}(\hat{s}_n < s_n(s))$. Now, $s_n(s)$ is equal to some s_k (using the labeling among the points of the grid). We have

$$\begin{aligned}
\mathbb{P}_{f,s}(\hat{s}_n < s_k) &= \sum_{j=1}^{k-1} \mathbb{P}_{f,s}(\hat{s}_n = s_j) \\
&\leq \sum_{j=1}^{k-1} \mathbb{P}_{f,s} \left(|\hat{\Phi}_n^p(u_n)| \geq \frac{1}{2} \left\{ q_{\beta'} \Phi^{[j]} + \Phi^{[j+1]} \right\} (u_n) \right) \\
&\leq \sum_{j=1}^{k-1} \mathbb{P}_{f,s} \left(|\hat{\Phi}_n^p(u_n) - \Phi^p(u_n)| \geq \frac{1}{2} \left\{ q_{\beta'} \Phi^{[j]} + \Phi^{[j+1]} \right\} (u_n) - |\Phi^p(u_n)| \right).
\end{aligned}$$

As $|\Phi^p(u_n)| \geq q_{\beta'}(u_n) \Phi^g(u_n)$ for large enough n, we get

$$
\begin{aligned}
\mathbb{P}_{f,s}(\hat{s}_n < s_k) &\leq \sum_{j=1}^{k-1} \mathbb{P}_{f,s}\left(|\hat{\Phi}_n^p(u_n) - \Phi^p(u_n)| \geq \frac{1}{2}\left\{ q_{\beta'}\Phi^{[j]} + \Phi^{[j+1]} \right\}(u_n)(1+o(1)) \right) \\
&\leq \sum_{j=1}^{k-1} \exp\left[-\frac{n}{4}\left(A^2 u_n^{-2\beta'} \exp(-2u_n^{s_j}) + \exp(-2u_n^{s_{j+1}}) \right) \right] \\
&\leq N \exp\left(-\frac{n}{4} \exp(-2u_n^s) \right).
\end{aligned}
$$

Now consider the case $\hat{s}_n > s_k$.

$$
\begin{aligned}
&\mathbb{P}_{f,s}(\hat{s}_n > s_k) \leq \sum_{j=k+1}^{N} \mathbb{P}_{f,s}\left(|\hat{\Phi}_n^p(u_n)| \leq \frac{1}{2}\{q_{\beta'}\Phi^{[j-1]} + \Phi^{[j]}\}(u_n) \right) \\
&\leq \sum_{j=k+1}^{N} \mathbb{P}_{f,s}\left(|\hat{\Phi}_n^p(u_n) - \Phi^p(u_n)| \geq q_{\beta'}(u_n)\Phi^g(u_n) - \frac{1}{2}\{q_{\beta'}\Phi^{[j-1]} + \Phi^{[j]}\}(u_n) \right) \\
&\leq N\mathbb{P}_{f,s}\left(|\hat{\Phi}_n^p(u_n) - \Phi^p(u_n)| \geq q_{\beta'}(u_n)\{\Phi^g(u_n) - \frac{1}{2}\Phi^{[k]}(u_n)\} + o(q_{\beta'}(u_n)\Phi^g(u_n)) \right)
\end{aligned}
$$

as $|\Phi^p(u_n)| \geq q_{\beta'}(u_n)\Phi^g(u_n)$ for large enough n and $j-1 \geq k$. According to the choice of the grid, we have $|s - s_k| \leq d_n$ and $d_n \log u_n \to 0$, which implies

$$
\begin{aligned}
\Phi^g(u_n) - \frac{1}{2}\Phi^k(u_n) &= \exp(-u_n^s)\left(1 - \frac{1}{2}\exp[u_n^{s_k}(u_n^{s-s_k} - 1)] \right) \\
&= \exp(-u_n^s)\left(1 - \frac{1}{2}\exp[u_n^{s_k}(s - s_k)\log u_n(1+o(1))] \right) \\
&\geq \exp(-u_n^s)\left(1 - \frac{1}{2}\exp(u_n^{s_k - \bar{s}}(1+o(1))) \right) \\
&\geq \frac{1}{2}\exp(-u_n^s)(1+o(1)),
\end{aligned}
$$

where the first inequality comes from $d_n \log u_n \leq u_n^{-\bar{s}}$. This gives

$$
\begin{aligned}
\mathbb{P}_{f,s}(\hat{s}_n > s_k) &\leq N\mathbb{P}_{f,s}\left(|\hat{\Phi}_n^p(u_n) - \Phi^p(u_n)| \geq \frac{1}{2}q_{\beta'}(u_n)\Phi^g(u_n)(1+o(1)) \right) \\
&\leq N \exp\left(-\frac{A^2}{2} n u_n^{-2\beta'} \exp(-2u_n^s)(1+o(1)) \right).
\end{aligned}
$$

In conclusion, as soon as we have $d_n \log u_n \leq u_n^{-\bar{s}}$, and $\log N = o((\log n)^\alpha)$ (which is ensured by our choice of d_n) we get, for any $\epsilon > 0$ and large enough n,

$$
\begin{aligned}
\mathbb{P}_{f,s}(|\hat{s}_n - s| \geq \epsilon) &\leq N \exp\left(-\frac{A^2}{2} n u_n^{-2\beta'} \exp(-2u_n^s)(1+o(1)) \right) \\
&\leq \exp\left(-\frac{A^2}{2}(\log n)^\alpha (1+o(1)) \right).
\end{aligned}
$$

The last term gives a convergent series and then according to Borel Cantelli's lemma, $\mathbb{P}_{f,s}(|\hat{s}_n - s| \geq \epsilon \text{ i.o }) = 0$ leading to the almost sure convergence of $\hat{s}_n$. ■

Proof of Corollary 1. Note that the new choice of d_n still satisfies the requirements for Proposition 1 to be valid. We introduce respectively, h_n, the non-random version of the bandwidth $\hat{h}_n$ and K_n the non-random version of the kernel $\hat{K}_n$ both constructed with self-similarity index $s_n(s)$. The Fourier transform Φ^{K_n} of K_n thus satisfies

$$\Phi^{K_n}(u) = \exp((|u|/h_n)^{s_n(s)})1_{|u|\leq 1}$$
$$\text{where } h_n = (2^{-1}\log n - (\bar{\beta} - s_n(s) + 1/2)\log\log n/s_n(s))^{-1/s_n(s)}.$$

We also introduce the corresponding (classical) estimator

$$f_n(x) = (nh_n)^{-1}\sum_{i=1}^{n} K_n(h_n^{-1}(x - Y_i)).$$

Note that obviously, $s_n(s), K_n$ and h_n are unknown to the statistician. These objects are used only as tools to assess the convergence of the procedure. Now, remark that we have

$$\mathbb{E}_{f,s}[|\hat{f}_n(x)-f(x)|^2] = \mathbb{E}_{f,s}[|f_n(x)-f(x)|^2 1_{\hat{s}_n=s_n(s)}]+\mathbb{E}_{f,s}[|\hat{f}_n(x)-f(x)|^2 1_{\hat{s}_n\neq s_n(s)}]$$
$$= T_1 + T_2,$$

say. Let us focus on the first term

$$T_1 \leq \mathbb{E}_{f,s}[|f_n(x) - f(x)|^2] = \{\mathbb{E}_{f,s}[f_n(x)] - f(x)\}^2 + \mathbb{V}\text{ar}_s\{f_n(x)\},$$

introducing the bias and the variance of the estimator $f_n(x)$. The important thing to note is that the kernel estimator f_n uses parameter $s_n(s)$ which is not equal to the true one s. Thus T_1 is not the classical risk for kernel estimator with known index s. Using Parseval's equality

$$\{\mathbb{E}_{f,s}[f_n(x)]-f(x)\}^2 = \frac{1}{4\pi}\left[\int e^{-iux}\Phi(u)\left(1_{|u|\leq 1/h_n}\exp(-|u|^s + |u|^{s_n(s)}) - 1\right)du\right]^2$$
$$\leq \frac{1}{4\pi}\left[\int_{|u|\leq 1/h_n}|\Phi(u)|\left(\exp(-|u|^s + |u|^{s_n(s)}) - 1\right)du + \int_{|u|>1/h_n}|\Phi(u)|du\right]^2.$$

The second term in the right hand side is the classical bias and equals $O(h_n^{\beta-1/2})$. As soon as $d_n h_n^{-s}\log(1/h_n)$ converges to zero, we can use the following development in the first term, uniformly for $|u| \leq 1/h_n$,

$$
\begin{aligned}
\exp(-|u|^s + |u|^{s_n(s)}) - 1 &= \exp\{|u|^s(s_n(s) - s)\log|u|(1+o(1))\} - 1 \\
&= |u|^s(s_n(s) - s)\log|u|(1+o(1)),
\end{aligned}
$$

which leads to

$$
\begin{aligned}
&\{\mathbb{E}_{f,s}[f_n(x)] - f(x)\}^2 \\
&\leq \frac{1}{4\pi}\left[\left(\int_{|u|\leq 1/h_n} |\Phi(u)||u|^s(s_n(s) - s)\log|u|du\right)(1+o(1)) + O(h_n^{\beta-1/2})\right]^2 \\
&\leq O(d_n^2)1_{\beta>s+1/2} + O(d_n^2 h_n^{2\beta-2s-1}\log^2(1/h_n))1_{\beta\leq s+1/2} + O(h_n^{2\beta-1}).
\end{aligned}
$$

It can be easily seen that

$$
\begin{aligned}
&d_n^2 h_n^{-2s}\log^2(1/h_n) \\
&\leq \frac{O(1)(\log n)^{2s/s_n(s)}(\log\log n)^2}{\log^2 n(\log\log n)^2} = O(1)(\log n)^{2(s-s_n(s))/s_n(s)} \\
&\leq O(1)(\log n)^{2d_n/\underline{s}} \leq O(1)\exp\{2\bar{s}/(\underline{s}\log n)\} = O(1),
\end{aligned}
$$

leading to

$$
\{\mathbb{E}_{f,s}[f_n(x)] - f(x)\}^2 \leq O(d_n^2)1_{\beta>s+1/2} + O(h_n^{2\beta-1}).
$$

Moreover, when $\beta > s + 1/2$, we use $d_n^2 \leq (\log n)^{-(2\bar{\beta}-1)/\underline{s}} = O(h_n^{(2\beta-1)})$. With this choice of d_n, we thus ensure that in any case

$$
\{\mathbb{E}_{f,s}[f_n(x)] - f(x)\}^2 \leq O(h_n^{2\beta-1}).
$$

The variance of $f_n(x)$ is bounded by

$$
\begin{aligned}
\mathbb{V}\mathrm{ar}_{f,s}\{f_n(x)\} &= \frac{1}{4\pi^2 n}\mathbb{E}_{f,s}\left[\int_{|u|\leq 1/h_n} e^{-iux}e^{|u|^{s_n(s)}}(e^{iuY} - \Phi^p(u))du\right]^2 \\
&\leq \frac{1}{\pi^2 n}\left(\int_{|u|\leq 1/\bar{h}_n} e^{|u|^{s_n(s)}}du\right)^2 = O\left(\frac{h_n^{2(s_n(s)-1)}\exp(2/h_n^{s_n(s)})}{n}\right).
\end{aligned}
$$

We finally get the bound

$$
T_1 \leq O(h_n^{2\beta-1}) + O\left(\frac{h_n^{2(s_n(s)-1)}\exp(2/h_n^{s_n(s)})}{n}\right).
$$

Now, we prove that the second term T_2 is negligible in front of the main term T_1, by using Proposition 1 and uniform bounds on $|\hat{f}_n(x)|$ and $|f(x)|$. First,

$$
\begin{aligned}
|\hat f_n(x)| &\le \int e^{|t|^{\bar s}} 1_{|t|\le 1/\hat h_n} dt = O(\hat h_n^{\bar s -1} \exp\{1/\hat h_n^{\bar s}\})\\
&\le O(1)(\log n)^{(1-\bar s)/\underline{s}} \exp\{(\log n)^{\bar s/\underline{s}}\}\\
|f(x)| &\le \int |\Phi(t)| dt = O(\int (1+|t|^{2\beta})^{-1} dt) = O(1),
\end{aligned}
$$

and then

$$
\begin{aligned}
T_2 &= O((\log n)^{2(1-\bar s)/\underline{s}} \exp\{2(\log n)^{\bar s/\underline{s}}\}) \mathbb{P}_{f,s}(\hat s_n \neq s_n(s))\\
&= O\left((\log n)^{2(1-\bar s)/\underline{s}} \exp\left(2(\log n)^{\bar s/\underline{s}} - \frac{A^2}{4}(\log n)^{\alpha}(1+o(1))\right)\right).
\end{aligned}
$$

As soon as we choose $\alpha > \bar s/\underline{s}$, this second term T_2 will be negligible in front of T_1. In conclusion,

$$
\begin{aligned}
\mathbb{E}_{f,s}[|\hat f_n(x) - f(x)|^2] &= O(h_n^{2\beta-1}) + O\left(h_n^{2(s_n(s)-1)} \frac{\exp(2/h_n^{s_n(s)})}{n}\right)\\
&= O((\log n)^{-(2\beta-1)/s_n(s)}) = O((\log n)^{-(2\beta-1)/s}).
\end{aligned}
$$

■

References

[1] Butucea, C. (2004a). Asymptotic normality of the integrated square error of a density estimator in the convolution model. *SORT*, **28**(1), 9–26.

[2] Butucea, C. (2004b). Deconvolution of supersmooth densities with smooth noise. *Can. J. Stat.*, **32**(2), 181–192.

[3] Butucea, C. (2004c). Quadratic functional estimation in view of minimax goodness-of-fit testing from noisy data. *To appear in Ann. Stat.* Available at `arXiv:math.ST/0612361` (short version) and `http://www.proba.jussieu.fr/pageperso/butucea/RevAOSlong.pdf` (long version).

[4] Butucea, C. and Matias, C. (2005). Minimax estimation of the noise level and of the deconvolution density in a semiparametric convolution model. *Bernoulli*, **11**(2), 309–340.

[5] Butucea, C. and Tsybakov, A. B. (2007). Sharp optimality for density deconvolution with dominating bias. I. II. *Theory Probab. Appl*, **52**. To appear.

[6] Carroll, R. J. and Hall, P. (1988). Optimal rates of convergence for deconvolving a density. *J. Amer. Statist. Assoc.*, **83**(404), 1184–1186.

[7] Comte, F., Rozenholc, Y., and Taupin, M.-L. (2006). Penalized contrast estimator for adaptive density deconvolution. *Can. J. Stat.*, **34**(3), 431–452.

[8] Efromovich, S. (1997). Density estimation for the case of supersmooth measurement error. *J. Amer. Statist. Assoc.*, **92**(438), 526–535.

[9] Fan, J. (1991). On the optimal rates of convergence for nonparametric deconvolution problems. *Ann. Statist.*, **19**(3), 1257–1272.

[10] Fan, J. (1992). Deconvolution with supersmooth distributions. *Can. J. Stat.*, **20**(2), 155–169.

[11] Fan, J. and Koo, J.-Y. (2002). Wavelet deconvolution. *IEEE Trans. Inf. Theory*, **48**(3), 734–747.

[12] Gayraud, G. and Pouet, C. (2005). Adaptive minimax testing in the discrete regression scheme. *Probab. Theory Relat. Fields*, **133**(4), 531–558.

[13] Hall, P. (1984). Central limit theorem for integrated square error of multivariate nonparametric density estimators. *J. Multivar. Anal.*, **14**(3), 1–16.

[14] Hall, P. and Heyde, C. (1980). *Martingale limit theory and its application*. Probability and Mathematical Statistics. Academic Press, New York.

[15] Holzmann, H., Bissantz, N., and Munk, A. (2007). Density testing in a contaminated sample. *J. Multivariate Analysis*, **98**(1), 57–75.

[16] Ingster, Y. and Suslina, I. (2003). *Nonparametric goodness-of-fit testing under Gaussian models*. Lecture Notes in Statistics. 169. Springer., New York.

[17] Korolyuk, V. and Borovskikh, Y. (1994). *Theory of U-statistics*. Mathematics and its Applications. Kluwer Academic Publishers, Dordrecht.

[18] Lehmann, E. L. and Romano, J. P. (2005). *Testing statistical hypotheses*. Springer Texts in Statistics. Springer, New York, third edition.

[19] Matias, C. (2002). Semiparametric deconvolution with unknown noise variance. *ESAIM, Probab. Stat.*, **6**, 271–292. (electronic).

[20] Meister, A. (2004). On the effect of misspecifying the error density in a deconvolution problem. *Can. J. Stat.*, **32**(4), 439–449.

[21] Pensky, M. and Vidakovic, B. (1999). Adaptive wavelet estimator for nonparametric density deconvolution. *Ann. Statist.*, **27**(6), 2033–2053.

A Technical Proofs

Proof of Lemma 2. Using a Markov inequality and the usual controls on bias and variance, we get

$$\mathbb{P}_0(|T_{n,N+1} - \mathbb{E}_0(T_{n,N+1})| > \mathcal{C}^\star t^2_{n,N+1} - \mathbb{E}_0(T_{n,N+1})) \leq \frac{Cn^{-2}(h^{N+1})^{-(4\sigma+1)}}{(\mathcal{C}^\star t^2_{n,N+1} - C(h^{N+1})^{2\bar{\beta}})^2} = O(\frac{1}{\mathcal{C}^\star - C}),$$

and by choosing $\mathcal{C}^\star$ large enough, this term is smaller than some $\epsilon > 0$. ■

Proof of Lemma 4. Let us write

$$\mathbb{P}_f(|T_{n,N+1}| \leq \mathcal{C}^\star t^2_{n,N+1}) \leq \mathbb{P}_f(|T_{n,N+1} - \mathbb{E}_f T_{n,N+1}| \geq \|f - f_0\|_2^2 - \mathcal{C}^\star t^2_{n,N+1} - B_f(T_{n,N+1}))$$

where

$$\begin{aligned}
&|B_f(T_{n,N+1})| = |\mathbb{E}_f(T_{n,N+1}) - \|f - f_0\|_2^2| \\
&\leq \int_{|u|\geq 1/h^{N+1}} |\Phi(u)|^2 du + 2\left(\int_{|u|\geq 1/h^{N+1}} |\Phi(u)|^2 du \int_{|u|\geq 1/h^{N+1}} |\Phi_0(u)|^2 du\right)^{1/2} \\
&\leq \Big(L(h^{N+1})^{2\beta}\exp\{-2\alpha/(h^{N+1})^r\} + 2L(h^{N+1})^{\beta+\bar{\beta}}\exp\{-\alpha/(h^{N+1})^r\}\Big)(1 + o(1)) \\
&\leq 2L(h^{N+1})^{\beta+\bar{\beta}}\exp\{-\alpha/(h^{N+1})^r\}(1 + o(1)).
\end{aligned}$$

In the same way as in the proof of Lemma 3, we have

$$\mathbb{E}_f(T_{n,N+1} - \mathbb{E}_f(T_{n,N+1}))^2 \leq \frac{C}{n^2(h^{N+1})^{4\sigma+1}} + \frac{4\Omega_g^2(f - f_0)}{n} 1_{\beta \geq \sigma} = w^2_{n,f},$$

and $\Omega_g(f - f_0)$ is a constant depending on f and g (but not n) and satisfying $|\Omega_g^2(f - f_0)| \leq C\|f - f_0\|_2^{2-2\sigma/\bar{\beta}}$. The rest of the proof follows the same lines as Lemma 3. Indeed, Markov's Inequality leads the following bound on the second type error term

$$\begin{aligned}
&\frac{w^2_{n,f}}{(\|f - f_0\|_2^2 - \mathcal{C}^\star t^2_{n,N+1} - 2L(h^{N+1})^{2\beta}\exp\{-\alpha/(h^{N+1})^r\}(1 + o(1)))^2} \\
&\qquad \leq \max\left(\frac{Cn^{-2}(h^{N+1})^{-4\sigma-1}}{(\mathcal{C}^0 - \mathcal{C}^\star)^2\psi^4_{n,r}}; \frac{C}{n\|f - f_0\|_2^{2+2\sigma/\bar{\beta}}(\mathcal{C}^0 - \mathcal{C}^\star)^2}\right)
\end{aligned}$$

The first term in the right hand side is a constant which can be as small as we need, by choosing a large enough constant $\mathcal{C}^0$. The second term converges to zero. ■

Proof of Lemma 5. As $\beta > \nu$, the bandwidths satisfy $h_\nu h_\beta^{-1} = o(1)$. Then, as G is compactly supported on $[-1,0]$, we have

$$\mathbb{E}_0\left(\frac{G_\beta\left(Y_1-x_{j,\beta}\right)G_\nu\left(Y_1-x_{i,\nu}\right)}{p_0^2\left(Y_1\right)}\right) = \int_{\mathbb{R}} \frac{G_\beta\left(y-x_{j,\beta}\right)G_\nu\left(y-x_{i,\nu}\right)}{p_0\left(y\right)}\,dy$$
$$= \int_{[-1,\,0]} \frac{G_\beta\left(h_\nu u + x_{i,\nu} - x_{j,\beta}\right)G\left(u\right)}{p_0\left(h_\nu u + x_{i,\nu}\right)}\,du.$$

Apply the Taylor Formula to get

$$G_\beta\left(h_\nu u + x_{i,\nu} - x_{j,\beta}\right) = G_\beta\left(x_{i,\nu} - x_{j,\beta}\right) + \frac{h_\nu}{h_\beta^2} u G'\left(\frac{h_\nu \tilde{u}_1 + x_{i,\nu} - x_{j,\beta}}{h_\beta}\right)$$
$$\text{and} \quad \frac{1}{p_0\left(h_\nu u + x_{i,\nu}\right)} = \frac{1}{p_0\left(x_{i,\nu}\right)} - \frac{p_0'\left(h_\nu \tilde{u}_2 + x_{i,\nu}\right)}{p_0\left(h_\nu \tilde{u}_2 + x_{i,\nu}\right)^2} h_\nu u,$$

where $0 \le \tilde{u}_1 \le u$ and $0 \le \tilde{u}_2 \le u$. As $\int G = 0$, we obtain

$$\int_{[-1,0]} \frac{G_\beta\left(h_\nu u + x_{i,\nu} - x_{j,\beta}\right)G\left(u\right)}{p_0\left(h_\nu u + x_{i,\nu}\right)}\,du$$
$$= \frac{1}{p_0\left(x_{i,\nu}\right)}\frac{h_\nu}{h_\beta^2}\int_{[-1,0]} uG'\left(\frac{h_\nu \tilde{u}_1 + x_{i,\nu} - x_{j,\beta}}{h_\beta}\right)G(u)du$$
$$-h_\nu G_\beta\left(x_{i,\nu} - x_{j,\beta}\right)\int_{[-1,0]} \frac{p_0'\left(h_\nu \tilde{u}_2 + x_{i,\nu}\right)}{p_0\left(h_\nu \tilde{u}_2 + x_{i,\nu}\right)^2} uG\left(u\right)du$$
$$-\frac{h_\nu^2}{h_\beta^2}\int_{[-1,0]} \frac{p_0'\left(h_\nu \tilde{u}_2 + x_{i,\nu}\right)}{p_0\left(h_\nu \tilde{u}_2 + x_{i,\nu}\right)^2} u^2 G'\left(\frac{h_\nu \tilde{u}_1 + x_{i,\nu} - x_{j,\beta}}{h_\beta}\right)G\left(u\right)\,du.$$

This leads to

$$\mathbb{E}_0\left(\frac{G_\beta\left(Y_1-x_{j,\beta}\right)G_\nu\left(Y_1-x_{i,\nu}\right)}{p_0^2\left(Y_1\right)}\right) = \frac{h_\nu}{h_\beta^2} R_{i,j}$$

where

$$R_{i,j} = \frac{1}{p_0\left(x_{i,\nu}\right)}\int_{[-1,0]} uG'\left(\frac{h_\nu \tilde{u}_1 + x_{i,\nu} - x_{j,\beta}}{h_\beta}\right)G(u)du$$
$$-h_\beta G\left(\frac{x_{i,\nu} - x_{j,\beta}}{h_\beta}\right)\int_{[-1,0]} \frac{p_0'\left(h_\nu \tilde{u}_2 + x_{i,\nu}\right)}{p_0\left(h_\nu \tilde{u}_2 + x_{i,\nu}\right)^2} uG\left(u\right)du$$
$$-h_\nu \int_{[-1,0]} \frac{p_0'\left(h_\nu \tilde{u}_2 + x_{i,\nu}\right)}{p_0\left(h_\nu \tilde{u}_2 + x_{i,\nu}\right)^2} u^2 G'\left(\frac{h_\nu \tilde{u}_1 + x_{i,\nu} - x_{j,\beta}}{h_\beta}\right)G\left(u\right)\,du.$$

satisfies

$$|R_{ij}| \le (\inf_{[0,1]} p_0)^{-1}\|G\|_\infty\|G'\|_\infty + \|G\|_\infty\|p_0'\|_\infty(\inf_{[-1,1]} p_0)^{-2}(h_\beta\|G\|_\infty + h_\nu\|G'\|_\infty),$$

which ends the proof of Lemma 5. ■

Proof of Theorem 2.

Assume now that $f_0 \in \mathcal{F}(\overline{\alpha}, \bar{r}, \beta_0, L)$, for some $\beta_0 \in [\underline{\beta}, \bar{\beta}]$. The proof follows the same lines as the proof of Theorem 1.

For the first-type error we write

$$\begin{aligned}\mathbb{P}_0(\Delta_n^* = 1) = &\sum_{i=1}^{N_1} \mathbb{P}_0(|T_{n,i} - \mathbb{E}_0(T_{n,i})| > \mathcal{C}^\star t_{n,i}^2 - \mathbb{E}_0(T_{n,i})) \\ &+ \sum_{i-N_1=1}^{N_2} \mathbb{P}_0(|T_{n,i} - \mathbb{E}_0(T_{n,i})| > \mathcal{C}^\star t_{n,i}^2 - \mathbb{E}_0(T_{n,i})).\end{aligned}$$

For the first N_1 terms we apply Lemma 1 with $\mathbb{E}_0(T_{n,i}) = o(1)L(h_i)^{2\beta_0}\exp(-2\overline{\alpha}/h_i^{\bar{r}})$ which is smaller than $t_{n,i}^2$ for all $i = 1, \dots, N_1$ and the same result follows. For the last N_2 terms we also use the Berry-Esseen inequality as in the proof of Lemma 1 for

$$x = \mathcal{C}^\star t_{n,i}^2 - \mathbb{E}_0(T_{n,i}) \ge \mathcal{C}^\star t_{n,i}^2(1 - o(1))$$

as $\mathbb{E}_0(T_{n,i}) = o(1)h_i^{2\beta_0}\exp(-2\overline{\alpha}/h_i^{\bar{r}}) = o(1/n)$. We get $x/v_n = O(1)(\log\log\log n)^{1/2}$

$$\begin{aligned}&\sum_{i-N_1=1}^{N_2} \mathbb{P}_0(|T_{n,i} - \mathbb{E}_0(T_{n,i})| > \mathcal{C}^\star t_{n,i}^2 - \mathbb{E}_0(T_{n,i})) \\ &\le N_2 \frac{v_n}{\mathcal{C}^\star t_{n,i}^2} \exp\left(-\frac{(\mathcal{C}^\star)^2 t_{n,i}^4}{4v_n^2}\right) \le C_1 \frac{(\log\log\log n)^{-1/2}}{(\log\log n)^{b-1}} = o(1),\end{aligned}$$

for some $b > 1$ for $\mathcal{C}^\star$ large enough. Indeed, all other calculations are similar as they are related mostly to the distribution of the noise which didn't change.

As for the second-type error,

$$\begin{aligned}&\sup_{\tau\in\mathcal{T}} \sup_{f\in\mathcal{F}(\tau,L)} \mathbb{P}_f(\Delta_n^\star = 0) \\ &\le 1_{\underline{r}=\bar{r}=0} \sup_{\alpha\ge\underline{\alpha},\beta\in[\underline{\beta};\bar{\beta}]} \sup_{\substack{f\in\mathcal{F}(\alpha,0,\beta,L)\\ \|f-f_0\|_2^2\ge\mathcal{C}\psi_{n,(\alpha,0,\beta)}^2}} \mathbb{P}_f(\forall 1 \le i \le N_1,\ |T_{n,i}| \le \mathcal{C}^\star t_{n,i}^2) \\ &+ 1_{\underline{r}>0} \sup_{r\in[\underline{r};\bar{r}],\alpha\in[\underline{\alpha},\overline{\alpha}],\beta\in[\underline{\beta},\bar{\beta}]} \sup_{\substack{f\in\mathcal{F}(\tau,L)\\ \|f-f_0\|_2^2\ge\mathcal{C}\psi_{n,\tau}^2}} \mathbb{P}_f(\forall N_1 + 1 \le i \le N_1 + N_2,\ |T_{n,i}| \le \mathcal{C}^\star t_{n,i}^2).\end{aligned}$$

For the first term in the previous sum we actually apply precisely Lemma 3. For the second term we mimic the proof of Lemma 3 and choose some f in $\mathcal{F}(\alpha, r, \beta, L)$ such that $\|f - f_0\|_2^2 \geq \mathcal{C}\psi_{n,r}^2$, where we denote $\psi_{n,r} = \psi_{n,\tau} 1_{r>0}$. We define r_f as the smallest point on the grid $\{r_1, \ldots, r_{N_2}\}$ such that $r \leq r_f$. We denote by h_f, $t_{n,f}^2$ and $T_{n,f}$ the bandwidth, the threshold and the test statistic associated to parameters $\overline{\alpha}$ and r_f (they do not depend on β). Then

$$\begin{aligned} &\mathbb{P}_f(\forall N_1 + 1 \leq i \leq N_1 + N_2,\ |T_{n,i}| \leq \mathcal{C}^\star t_{n,i}^2) \\ \leq &\mathbb{P}_f(|T_{n,f} - \mathbb{E}_f(T_{n,f})| \geq \|f - f_0\|_2^2 - \mathcal{C}^\star t_{n,f}^2 - B_f(T_{n,f})), \end{aligned} \tag{A.1}$$

where, as in Theorem 1

$$\begin{aligned} |B_f(T_{n,f})| &= |\|J_h * f - f\|_2^2 + 2\langle f - J_h * f, f_0\rangle| \\ &\leq \left(Lh_f^{2\beta} \exp(-2\alpha/h_f^r) + 2Lh_f^{\beta+\beta_0} \exp(-\alpha/h_f^r - \overline{\alpha}/h_f^{\bar{r}})\right)(1 + o(1)) \\ &\leq L(h_f^{2\beta} + h_f^{\beta+\beta_0}) \exp(-2\alpha/h_f^r)(1 + o(1)) \\ &\leq L(h_f^{\beta+\beta\wedge\beta_0}) \exp(-2\alpha/h_f^r)(1 + o(1)). \end{aligned}$$

Using Markov's inequality, we get the following upper bound for (A.1)

$$\frac{\mathbb{V}\mathrm{ar}_f(T_{n,f})}{(\|f - f_0\|_2^2 - \mathcal{C}^\star t_{n,f}^2 - B_f(T_{n,f}))^2}. \tag{A.2}$$

The variance is bounded from above by

$$\mathbb{E}_f(T_{n,f} - \mathbb{E}_f(T_{n,f}))^2 \leq \frac{C}{n^2 h_f^{4\sigma+1}} + \frac{4\Omega_g^2(f - f_0)}{n}, \tag{A.3}$$

and similarly to [3] we show that $\Omega_g^2(f - f_0) \leq \|f - f_0\|_2^2 (\log \|f - f_0\|_2^{-2})^{2\sigma/r}$. We have

$$t_{n,f}^2 \psi_{n,r}^{-2} = (\log n)^{(4\sigma+1)(1/r_f - 1/r)/2} \leq 1,$$

and thus $\|f - f_0\|_2^2 - \mathcal{C}^\star t_{n,f}^2 \geq (\mathcal{C} - \mathcal{C}^\star)\psi_{n,r}^2$. Moreover,

$$\begin{aligned} B_f(T_{n,f})\psi_{n,r}^{-2} \leq C(\log\log\log n)^{-1/2} &(\log n)^{-(\beta+\beta\wedge\beta_0)/r_f - (4\sigma+1)/(2r)} \\ &\times \exp\left\{-2\alpha \left(\frac{\log n}{2c}\right)^{r/r_f} + \log n\right\}. \end{aligned}$$

The construction of the grid ensures that $-1/(\log\log n) \leq r - r_f \leq 0$ and thus

$$
\begin{aligned}
&\exp\left\{-2\alpha\left(\frac{\log n}{2c}\right)^{r/r_f}+\log n\right\}\\
&=\exp\left\{-\frac{\log n}{c}\left[\alpha\exp\left(\frac{r-r_f}{r_f}\log\log n(1+o(1))\right)-c\right]\right\}\\
&\leq\exp\left\{-\frac{\log n}{c}\left[\underline{\alpha}\exp\left(\frac{-1}{\underline{r}}(1+o(1))\right)-c\right]\right\}=O(1),
\end{aligned}
$$

as we chose the constant $c < \underline{\alpha}\exp(-1/\underline{r})$. Finally, we have $B_f(T_{n,f})\psi_{n,r}^{-2} = o(1)$. Let us come back to (A.2). We distinguish two cases whether the first or the second term in (A.3) is dominant. If the first term in the variance dominates, we have the following bound for (A.2)

$$
\frac{n^{-2}h_f^{-(4\sigma+1)}}{(\mathcal{C}-\mathcal{C}^\star)^2\psi_{n,\tau}^4} \leq \frac{C}{\log\log\log n} \to 0.
$$

On the other hand, if the second term in (A.3) is the larger one, the bound (A.2) writes

$$
\begin{aligned}
\frac{n^{-1}\|f-f_0\|_2^2(\log\|f-f_0\|_2^{-2})^{2\sigma/r}}{\|f-f_0\|_2^4(1-\mathcal{C}^\star/\mathcal{C}+o(1))^2} &\leq Cn^{-1}\psi_{n,r}^{-2}(\log\psi_{n,r}^{-2})^{2\sigma/r}\\
&= C(\log n)^{-1/(2r)}(\log\log\log n)^{-1/2}=o(1).
\end{aligned}
$$

This finishes the proof. ■

Proof of Corollary 2. We keep on with the same notation as in Subsection 3.2 and denote by I the functional $\int f^2$. In the same way as in the proof of Corollary 1, we write

$$
\mathbb{E}_{f,s}[|\hat{T}_n - I|^2] \leq \mathbb{E}_{f,s}[|T_n - I|^2] + \mathbb{E}_{f,s}[|\hat{T}_n - I|^2 1_{\{\hat{s}_n \neq s_n(s)\}}]. \tag{A.4}
$$

Let us first focus on the first term appearing in the right hand side of (A.4). We split it into the square of a bias term plus a variance term. The bias is bounded by

$$
\begin{aligned}
|\mathbb{E}_{f,s}T_n - I| &\leq \frac{1}{2\pi}\left(\int_{|u|>1/h_n}|\Phi(u)|^2du + \int_{|u|\leq 1/h_n}|\exp(2|u|^{s_n(s)}-2|u|^s)-1||\Phi(u)|^2du\right)\\
&\leq O(h_n^{-2\beta}) + \int_{|u|\leq 1/h_n} 2|u|^s|s_n(s)-s|\log|u||\Phi(u)|^2du\\
&\leq O(h_n^{-2\beta}) + O(d_n)1_{\{s\leq 2\beta\}} + O(h_n^{2\beta-s}\log(1/h_n)d_n)1_{\{s>2\beta\}}.
\end{aligned}
$$

Like in the proof of Corollary 1, we have $d_n h_n^{-s}\log(1/h_n) = o(1)$ and thus using that $d_n \leq (\log n)^{-2\bar{\beta}/\underline{s}} = O((\log n)^{-2\beta/s})$, we finally get

$$
|\mathbb{E}_{f,s}T_n - I| \leq O((\log n)^{-2\beta/s}).
$$

Concerning the variance term, we easily get

$$\mathbb{V}\mathrm{ar}_{f,s}(T_n) \leq \frac{C_1}{n^2} h_n^{s_n(s)-1} \exp(4/h_n^{s_n(s)}) + \frac{C_2}{n} h_n^{2\beta+s_n(s)-1} \exp(2/h_n^{s_n(s)}),$$

where C_1 and C_2 are positive constants (we refer to [3], Theorem 4 for more details). Using the form of the bandwidth h_n, we have

$$\mathbb{E}_{f,s}|T_n - I|^2 = O\left(\frac{\log n}{2}\right)^{-4\beta/s}.$$

Let us now focus on the second term appearing in the right hand side of (A.4). Denoting by $h_0 = (\log n/2)^{-1/\underline{s}}$, we have

$$|\hat{T}_n| \leq \frac{1}{2\pi} \int_{|u|\leq 1/h_0} \exp(2|u|^{\bar{s}}) du = O(h_0^{\bar{s}-1} \exp(2/h_0^{\bar{s}})).$$

Moreover,

$$I = \|f\|_2^2 = \frac{1}{2\pi} \|\Phi\|_2^2$$

This leads to

$$\begin{aligned}
\mathbb{E}_{f,s}[|\hat{T}_n - I|^2 1_{\{\hat{s}_n \neq s_n(s)\}}] &\leq C\left(\frac{\log n}{2}\right)^{(1-\bar{s})/\underline{s}} \exp\left\{2\left(\frac{\log n}{2}\right)^{\bar{s}/\underline{s}}\right\} \mathbb{P}_{f,s}(\hat{s}_n \neq s_n(s)) \\
&\leq C\left(\frac{\log n}{2}\right)^{(1-\bar{s})/\underline{s}} \exp\left\{2\left(\frac{\log n}{2}\right)^{\bar{s}/\underline{s}}\right\} \exp\left(-\frac{A^2}{4}(\log n)^a (1+o(1))\right),
\end{aligned}$$

and this term is negligible in front of the first term appearing in the right hand side of (A.4) as soon as $a > \bar{s}/\underline{s}$. This leads to the result. ■

Proof of Corollary 3. We use the same notation as in Subsection 3.2. Moreover, T_n^0 is the test statistic constructed with the deterministic kernel K_n and the deterministic bandwidth h_n; and t_n^2 is the threshold defined with the parameter value $s_n(s)$ for the self-similarity index. The first type error of the test is controlled by

$$\mathbb{P}_{f_0,s}(\Delta_n^\star = 1) = \mathbb{P}_{f_0,s}(|\hat{T}_n^0| \hat{t}_n^{-2} > \mathcal{C}^\star) \leq \mathbb{P}_{f_0,s}(\hat{s}_n \neq s_n(s)) + \mathbb{P}_{f_0,s}(|T_n^0| t_n^{-2} > \mathcal{C}^\star).$$

The first term on the right hand side of this inequality converges to zero according to Proposition 1. Let us focus on the second term. We have

$$\mathbb{P}_{f_0,s}(|T_n^0| t_n^{-2} > \mathcal{C}^\star) \leq \frac{1}{(\mathcal{C}^\star)^2 t_n^4} \mathbb{E}_{f_0,s}(T_n^0)^2 \leq \frac{1}{(\mathcal{C}^\star)^2 t_n^4} \left\{ (\mathbb{E}_{f_0,s} T_n^0)^2 + \mathbb{V}\mathrm{ar}_{f_0,s} T_n^0 \right\}.$$

It is easily seen that

$$
\begin{aligned}
\mathbb{E}_{f_0,s}T_n^0 &= \frac{1}{2\pi}\int_{|u|\le 1/h_n}|\Phi_0(u)|^2|\exp(|u|^{s_n(s)}-|u|^s)-1|^2du+\frac{1}{2\pi}\int_{|u|>1/h_n}|\Phi_0(u)|^2du\\
&\le \frac{d_n^2}{2\pi}\left(\int_{|u|\le 1/h_n}|\Phi_0(u)|^2|u|^{2s}\log^2|u|du\right)(1+o(1))+O(h_n^{2\bar\beta})\\
&\le O(d_n^2)1_{\bar\beta>s}+O(h_n^{2\bar\beta})=O(h_n^{2\bar\beta}).
\end{aligned}
$$

the inequalities being valid as soon as $h_n^{-s}d_n\log(1/h_n)$ converges to zero. Like in the proof of Theorem 4 in [3], we can show that

$$
\mathbb{V}\mathrm{ar}_{f_0,s}(T_n^0)\le O(1)\frac{h_n^{s_n(s)-1}}{n^2}\exp(4/h_n^{s_n(s)})+O(1)\frac{h_n^{2\bar\beta+s_n(s)-1}}{n}\exp(2/h_n^{s_n(s)}).
$$

Finally, we get

$$
\begin{aligned}
&\mathbb{P}_{f_0,s}(|T_n^0|t_n^{-2}>\mathcal{C}^\star)\\
&\le \frac{1}{(\mathcal{C}^\star)^2t_n^4}\left\{O(h_n^{4\bar\beta})+O(1)\frac{h_n^{s_n(s)-1}}{n^2}\exp(4/h_n^{s_n(s)})+O(1)\frac{h_n^{2\bar\beta+s_n(s)-1}}{n}\exp(2/h_n^{s_n(s)})\right\}\\
&\le \frac{O(1)}{\mathcal{C}^\star}.
\end{aligned}
$$

Choosing $\mathcal{C}^\star$ large enough achieves the control of the first error term. We now turn to the second error term. Under hypothesis $H_1(\mathcal{C},\Psi_n)$, there exists some β such that f belongs to $\mathcal{F}(0,0,\beta,L)$ and $\|f-f_0\|_2^2\ge \mathcal{C}\psi_{n,\beta}$. We write

$$
\mathbb{P}_{f,s}(\Delta_n^\star=0)=\mathbb{P}_{f,s}(|\hat T_n^0|\hat t_n^{-2}\le \mathcal{C}^\star)\le \mathbb{P}_{f,s}(\hat s_n\ne s_n(s))+\mathbb{P}_{f,s}(|T_n^0|t_n^{-2}\le\mathcal{C}^\star).
$$

As already seen, the first term in the right hand side of this inequality converges to zero, so we only deal with the second one. We define $B_{f,s}(T_n^0)=\mathbb{E}_{f,s}T_n^0-\|f-f_0\|_2^2$. Thus

$$
\begin{aligned}
\mathbb{P}_{f,s}(|T_n^0|t_n^{-2}\le\mathcal{C}^\star)&\le \mathbb{P}_{f,s}(|T_n^0-\mathbb{E}_{f,s}T_n^0|\ge \|f-f_0\|_2^2-\mathcal{C}^\star t_n^2+B_{f,s}(T_n^0))\\
&\le \frac{\mathbb{V}\mathrm{ar}_{f,s}(T_n^0)}{(\|f-f_0\|_2^2-\mathcal{C}^\star t_n^2+B_{f,s}(T_n^0))^2}. \quad (A.5)
\end{aligned}
$$

We compute this bias term $B_{f,s}(T_n^0)$.

$$
\begin{aligned}
B_{f,s}(T_n^0) &= \frac{1}{2\pi}\int|\exp(|u|^{s_n(s)}-|u|^s)\Phi(u)1_{|u|\le 1/h_n}-\Phi_0(u)|^2du-\frac{1}{2\pi}\int|\Phi(u)-\Phi_0(u)|^2du\\
&\le \frac{1}{2\pi}\int_{|u|\le 1/h_n}|[\exp(|u|^{s_n(s)}-|u|^s)-1]\Phi(u)|^2du+\frac{1}{2\pi}\int_{|u|>1/h_n}|\Phi(u)|^2du\\
&\le \frac{d_n^2}{2\pi}(1+o(1))\int_{|u|\le 1/h_n}|u|^{2s}\log^2|u||\Phi(u)|^2du+O(h_n^{2\beta})\\
&\le O(d_n^2)1_{\beta>s}+O(h_n^{2\beta})=O(h_n^{2\beta}).
\end{aligned}
$$

In fact, there exists some constant $C_1 > 0$ depending only on L and on the noise distribution such that $B_{f,s}(T_n^0) \leq C_1 h_n^{2\beta}$. Under hypothesis $H_1(\mathcal{C}, \Psi_n)$, we also have $\|f - f_0\|_2^2 \geq \mathcal{C}\psi_{n,\beta}^2$. Thus,

$$\begin{aligned}\|f - f_0\|_2^2 - \mathcal{C}^\star t_n^2 + B_{f,s}(T_n^0) &\geq \mathcal{C}\left(\frac{\log n}{2}\right)^{-2\beta/s} - \mathcal{C}^\star\left(\frac{\log n}{2}\right)^{-2\bar{\beta}/s_n(s)} - C_1\left(\frac{\log n}{2}\right)^{-2\beta/s_n(s)} \\ &\geq a\left(\frac{\log n}{2}\right)^{-2\beta/s}.\end{aligned}$$

where $a = \mathcal{C} - \mathcal{C}^\star - C_1$ is positive whenever $\mathcal{C} > \mathcal{C}^0 := \mathcal{C}^\star - C_1$. Returning to (A.5), we get

$$\mathbb{P}_{f,s}(|T_n^0| t_n^{-2}) \leq \frac{\psi_{n,\beta}^4}{a^2}\mathbb{V}\mathrm{ar}_{f,s}(T_n^0).$$

Computation of the variance follows the same lines as under hypothesis H_0. We obtain

$$\mathbb{V}\mathrm{ar}_{f,s}(T_n^0) \leq O(1)\frac{h_n^{s_n(s)-1}}{n}\exp(2/h_n^{s_n(s)})\left(h_n^{2\beta} + \frac{\exp(2/h_n^{s_n(s)})}{n}\right).$$

The choice of the bandwidth ensures that the second type error term converges to zero. ■

Proof of Theorem 3. This proof follows the lines of Theorem 3.9 in [14]. Combining the Skorokhod representation Theorem and Lemma 3.3 in [14], there exists a nonnegative random variable T_n such that for any $0 < \epsilon < 1/2$ and any real x,

$$|\mathbb{P}(U_n \leq x) - \phi(x)| = |\mathbb{P}(S_n \leq v_n^{-1}x) - \phi(x/v_n)| \leq 16\epsilon^{1/2}\exp\{-x^2/(4v_n^2)\} + \mathbb{P}(|T_n - 1| > \epsilon).$$

Moreover, for any $\delta > 0$,

$$\mathbb{P}(|T_n - 1| > \epsilon) \leq 4\epsilon^{-1-\delta}\mathbb{E}\left[|T_n - V_n^2|^{1+\delta} + |V_n^2 - 1|^{1+\delta}\right],$$

where $T_n - V_n^2$ is a sum of Martingale differences. In the same way as in [14], we obtain (as $\delta \leq 1$)

$$\mathbb{P}(|T_n - 1| > \epsilon) \leq C\epsilon^{-1-\delta}\left[\sum_{i=1}^n \mathbb{E}|Z_i|^{2+2\delta} + \mathbb{E}|V_n^2 - 1|^{1+\delta}\right],$$

which concludes the proof. ■